(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

B4

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 September 2001 (13.09.2001)

PCT

(10) International Publication Number
WO 01/66742 A2

(51) International Patent Classification[7]: C12N 15/12, 15/10, 15/66, C07K 14/705, 16/28, A01K 67/027, A61K 38/17, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/US01/06814

(22) International Filing Date: 1 March 2001 (01.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/186,854	3 March 2000 (03.03.2000)	US
60/188,384	10 March 2000 (10.03.2000)	US
60/190,453	17 March 2000 (17.03.2000)	US
60/190,730	20 March 2000 (20.03.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94025 (US). **SHIH, Leo, L.** [US/US]; 1081 Tanland Drive, Apt. B, Palo Alto, CA 94303 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **KHAN, Farrah, A.** [IN/US]; 3617 Central Road #102, Glenview, Illinois 60025 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **HARLAND, Lee** [GB/GB]; 18 Chaucer Court, New Dover Road, Canterbury Kent CT1 3AU (GB). **WALSH, Roderick, T.** [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury Kent CT1 3EZ (GB). **LO, Terence, P.** [CA/US]; 1451 Beach Park Boulevard, Apt. 115, Foster City, CA 94404-1941 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/66742 A2

(54) Title: G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The invention provides human G-protein coupled receptors (GCREC) and polynucleotides which identify and encode GCREC. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of GCREC.

G-PROTEIN COUPLED RECEPTORS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of G-protein coupled receptors and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

BACKGROUND OF THE INVENTION

Signal transduction is the general process by which cells respond to extracellular signals. Signal transduction across the plasma membrane begins with the binding of a signal molecule, e.g., a hormone, neurotransmitter, or growth factor, to a cell membrane receptor. The receptor, thus activated, triggers an intracellular biochemical cascade that ends with the activation of an intracellular target molecule, such as a transcription factor. This process of signal transduction regulates all types of cell functions including cell proliferation, differentiation, and gene transcription. The G-protein coupled receptors (GPCRs), encoded by one of the largest families of genes yet identified, play a central role in the transduction of extracellular signals across the plasma membrane. GPCRs have a proven history of being successful therapeutic targets.

GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which together form a bundle of antiparallel alpha (α) helices. GPCRs range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10; Coughlin, S.R. (1994) Curr. Opin. Cell Biol. 6:191-197). The amino-terminus of a GPCR is extracellular, is of variable length, and is often glycosylated. The carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops alternate with intracellular loops and link the transmembrane domains. Cysteine disulfide bridges linking the second and third extracellular loops may interact with agonists and antagonists. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account, in part, for structural and functional features of the receptor. In most cases, the bundle of α helices forms a ligand-binding pocket. The extracellular N-terminal segment, or one or more of the three extracellular loops, may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. In turn, the large, third intracellular loop of the activated receptor interacts with a heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, including the activation of second

messengers such as cyclic AMP (cAMP), phospholipase C, and inositol triphosphate, and the interaction of the activated GPCR with ion channel proteins. (See, e.g., Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6; Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego CA, pp. 162-176; Baldwin, J.M. (1994) Curr. Opin. Cell Biol. 6:180-190.)

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, γ-aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., bombesin, bradykinin, calcitonin, C5a anaphylatoxin, endothelin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors.

The diversity of the GPCR family is further increased by alternative splicing. Many GPCR genes contain introns, and there are currently over 30 such receptors for which splice variants have been identified. The largest number of variations are at the protein C-terminus. N-terminal and cytoplasmic loop variants are also frequent, while variants in the extracellular loops or transmembrane domains are less common. Some receptors have more than one site at which variance can occur. The splicing variants appear to be functionally distinct, based upon observed differences in distribution, signaling, coupling, regulation, and ligand binding profiles (Kilpatrick, G.J. et al. (1999) Trends Pharmacol. Sci. 20:294-301).

GPCRs can be divided into three major subfamilies: the rhodopsin-like, secretin-like, and metabotropic glutamate receptor subfamilies. Members of these GPCR subfamilies share similar functions and the characteristic seven transmembrane structure, but have divergent amino acid sequences. The largest family consists of the rhodopsin-like GPCRs, which transmit diverse extracellular signals including hormones, neurotransmitters, and light. Rhodopsin is a photosensitive GPCR found in animal retinas. In vertebrates, rhodopsin molecules are embedded in membranous stacks found in photoreceptor (rod) cells. Each rhodopsin molecule responds to a photon of light by triggering a decrease in cGMP levels which leads to the closure of plasma membrane sodium channels. In this manner, a visual signal is converted to a neural impulse. Other rhodopsin-like GPCRs are directly involved in responding to neurotransmitters. These GPCRs include the receptors for adrenaline (adrenergic receptors), acetylcholine (muscarinic receptors), adenosine, galanin, and glutamate (N-

methyl-D-aspartate/NMDA receptors). (Reviewed in Watson, S. and S. Arkinstall (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 7-9, 19-22, 32-35, 130-131, 214-216, 221-222; Habert-Ortoli, E. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9780-9783.)

The galanin receptors mediate the activity of the neuroendocrine peptide galanin, which inhibits secretion of insulin, acetylcholine, serotonin and noradrenaline, and stimulates prolactin and growth hormone release. Galanin receptors are involved in feeding disorders, pain, depression, and Alzheimer's disease (Kask, K. et al. (1997) Life Sci. 60:1523-1533). Other nervous system rhodopsin-like GPCRs include a growing family of receptors for lysophosphatidic acid and other lysophospholipids, which appear to have roles in development and neuropathology (Chun, J. et al. (1999) Cell Biochem. Biophys. 30:213-242).

The largest subfamily of GPCRs, the olfactory receptors, are also members of the rhodopsin-like GPCR family. These receptors function by transducing odorant signals. Numerous distinct olfactory receptors are required to distinguish different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors are found in nasal passages. For example, the RA1c receptor which was isolated from a rat brain library, has been shown to be limited in expression to very distinct regions of the brain and a defined zone of the olfactory epithelium (Raming, K. et al. (1998) Receptors Channels 6:141-151). However, the expression of olfactory-like receptors is not confined to olfactory tissues. For example, three rat genes encoding olfactory-like receptors having typical GPCR characteristics showed expression patterns not only in taste and olfactory tissue, but also in male reproductive tissue (Thomas, M.B. et al. (1996) Gene 178:1-5).

Members of the secretin-like GPCR subfamily have as their ligands peptide hormones such as secretin, calcitonin, glucagon, growth hormone-releasing hormone, parathyroid hormone, and vasoactive intestinal peptide. For example, the secretin receptor responds to secretin, a peptide hormone that stimulates the secretion of enzymes and ions in the pancreas and small intestine (Watson, supra, pp. 278-283). Secretin receptors are about 450 amino acids in length and are found in the plasma membrane of gastrointestinal cells. Binding of secretin to its receptor stimulates the production of cAMP.

Examples of secretin-like GPCRs implicated in inflammation and the immune response include the EGF module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors subfamily. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes and is markedly upregulated on activated B and T cells (McKnight, A.J. and S. Gordon

(1998) J. Leukoc. Biol. 63:271-280).

The third GPCR subfamily is the metabotropic glutamate receptor family. Glutamate is the major excitatory neurotransmitter in the central nervous system. The metabotropic glutamate receptors modulate the activity of intracellular effectors, and are involved in long-term potentiation (Watson, supra, p.130). The Ca^{2+}-sensing receptor, which senses changes in the extracellular concentration of calcium ions, has a large extracellular domain including clusters of acidic amino acids which may be involved in calcium binding. The metabotropic glutamate receptor family also includes pheromone receptors, the $GABA_B$ receptors, and the taste receptors.

Other subfamilies of GPCRs include two groups of chemoreceptor genes found in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae, which are distantly related to the mammalian olfactory receptor genes. The yeast pheromone receptors STE2 and STE3, involved in the response to mating factors on the cell membrane, have their own seven-transmembrane signature, as do the cAMP receptors from the slime mold Dictyostelium discoideum, which are thought to regulate the aggregation of individual cells and control the expression of numerous developmentally-regulated genes.

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, supra). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Furthermore, somatic activating mutations in the thyrotropin receptor have been reported to cause hyperfunctioning thyroid adenomas, suggesting that certain GPCRs susceptible to constitutive activation may behave as protooncogenes (Parma, J. et al. (1993) Nature 365:649-651). GPCR receptors for the following ligands also contain mutations associated with human disease: luteinizing hormone (precocious puberty); vasopressin V_2 (X-linked nephrogenic diabetes); glucagon (diabetes and hypertension); calcium (hyperparathyroidism, hypocalcuria, hypercalcemia); parathyroid hormone (short limbed dwarfism); β_3-adrenoceptor (obesity, non-insulin-dependent diabetes mellitus); growth hormone releasing hormone (dwarfism); and adrenocorticotropin (glucocorticoid deficiency) (Wilson, S. et al. (1998) Br. J. Pharmocol. 125:1387-1392; Stadel, J.M. et al. (1997) Trends Pharmacol. Sci. 18:430-437). GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure, and several cardiovascular disorders (Horn, F. and G. Vriend (1998) J. Mol. Med. 76:464-468).

In addition, within the past 20 years several hundred new drugs have been recognized that are directed towards activating or inhibiting GPCRs. The therapeutic targets of these drugs span a wide range of diseases and disorders, including cardiovascular, gastrointestinal, and central nervous system disorders as well as cancer, osteoporosis and endometriosis (Wilson, supra; Stadel, supra). For example, the dopamine agonist L-dopa is used to treat Parkinson's disease, while a dopamine antagonist is used to treat schizophrenia and the early stages of Huntington's disease. Agonists and antagonists of

adrenoceptors have been used for the treatment of asthma, high blood pressure, other cardiovascular disorders, and anxiety; muscarinic agonists are used in the treatment of glaucoma and tachycardia; serotonin 5HT1D antagonists are used against migraine; and histamine H1 antagonists are used against allergic and anaphylactic reactions, hay fever, itching, and motion sickness (Horn, supra).

Recent research suggests potential future therapeutic uses for GPCRs in the treatment of metabolic disorders including diabetes, obesity, and osteoporosis. For example, mutant V2 vasopressin receptors causing nephrogenic diabetes could be functionally rescued in vitro by co-expression of a C-terminal V2 receptor peptide spanning the region containing the mutations. This result suggests a possible novel strategy for disease treatment (Schöneberg, T. et al. (1996) EMBO J. 15:1283-1291). Mutations in melanocortin-4 receptor (MC4R) are implicated in human weight regulation and obesity. As with the vasopressin V2 receptor mutants, these MC4R mutants are defective in trafficking to the plasma membrane (Ho, G. and R.G. MacKenzie (1999) J. Biol. Chem. 274:35816-35822), and thus might be treated with a similar strategy. The type 1 receptor for parathyroid hormone (PTH) is a GPCR that mediates the PTH-dependent regulation of calcium homeostasis in the bloodstream. Study of PTH/receptor interactions may enable the development of novel PTH receptor ligands for the treatment of osteoporosis (Mannstadt, M. et al. (1999) Am. J. Physiol. 277:F665-F675).

The chemokine receptor group of GPCRs have potential therapeutic utility in inflammation and infectious disease. (For review, see Locati, M. and P.M. Murphy (1999) Annu. Rev. Med. 50:425-440.) Chemokines are small polypeptides that act as intracellular signals in the regulation of leukocyte trafficking, hematopoiesis, and angiogenesis. Targeted disruption of various chemokine receptors in mice indicates that these receptors play roles in pathologic inflammation and in autoimmune disorders such as multiple sclerosis. Chemokine receptors are also exploited by infectious agents, including herpesviruses and the human immunodeficiency virus (HIV-1) to facilitate infection. A truncated version of chemokine receptor CCR5, which acts as a coreceptor for infection of T-cells by HIV-1, results in resistance to AIDS, suggesting that CCR5 antagonists could be useful in preventing the development of AIDS.

The discovery of new G-protein coupled receptors and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of G-protein coupled receptors.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, G-protein coupled receptors, referred to collectively as "GCREC" and individually as "GCREC-1," "GCREC-2," "GCREC-3," "GCREC-4," "GCREC-5," "GCREC-6," "GCREC-7," "GCREC-8," "GCREC-9," "GCREC-10," "GCREC-11," "GCREC-12," "GCREC-13," "GCREC-14," "GCREC-15," "GCREC-16," "GCREC-17," "GCREC-18," "GCREC-19," "GCREC-20," and "GCREC-21." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90%

sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as

an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the

activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted

motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"GCREC" refers to the amino acid sequences of substantially purified GCREC obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of GCREC. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

An "allelic variant" is an alternative form of the gene encoding GCREC. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding GCREC include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as GCREC or a polypeptide with at least one functional characteristic of GCREC. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding GCREC, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GCREC. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GCREC. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of GCREC is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known

in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of GCREC. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of GCREC either by directly interacting with GCREC or by acting on components of the biological pathway in which GCREC participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind GCREC polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"

refers to the capability of the natural, recombinant, or synthetic GCREC, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding GCREC or fragments of GCREC may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val

Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of GCREC or the polynucleotide encoding GCREC which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that

specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search

Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions,

explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of GCREC which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of GCREC which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of GCREC. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of GCREC.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an GCREC may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by

cell type depending on the enzymatic milieu of GCREC.

"Probe" refers to nucleic acid sequences encoding GCREC, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for

microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing GCREC, nucleic acids encoding GCREC, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a

certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human G-protein coupled receptors (GCREC), the polynucleotides encoding GCREC, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are G-protein coupled receptors. For example, SEQ ID NO:7 is 85% identical, from residue M1 to residue V306, to murine odorant receptor MOR83

(GenBank ID g6178006) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.5e-141, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a 7 transmembrane receptor (rhodopsin family) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is an olfactory G-protein coupled receptor. SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:22-42 or that distinguish between SEQ ID NO:22-42 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6871486H1 is the identification number of an Incyte cDNA sequence, and BRAGNON02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70171099V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5743982) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6671985_006 is the

identification number of a Genscan-predicted coding sequence, with g6671985 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL2289894_00001 represents a "stitched" sequence in which 2289894 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses GCREC variants. A preferred GCREC variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the GCREC amino acid sequence, and which contains at least one functional or structural characteristic of GCREC.

The invention also encompasses polynucleotides which encode GCREC. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42, which encodes GCREC. The polynucleotide sequences of SEQ ID NO:22-42, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding GCREC. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding GCREC. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:22-42 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:22-42.

Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of GCREC.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding GCREC, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring GCREC, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GCREC and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring GCREC under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GCREC or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GCREC and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode GCREC and GCREC derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GCREC or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding GCREC may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full'length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library

does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GCREC may be cloned in recombinant DNA molecules that direct expression of GCREC, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express GCREC.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GCREC-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of GCREC, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random

point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding GCREC may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, GCREC itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of GCREC, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active GCREC, the nucleotide sequences encoding GCREC or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding GCREC. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GCREC. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding GCREC and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be

enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GCREC and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GCREC. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding GCREC. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding GCREC can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding GCREC into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro

transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of GCREC are needed, e.g. for the production of antibodies, vectors which direct high level expression of GCREC may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of GCREC. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of GCREC. Transcription of sequences encoding GCREC may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GCREC may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses GCREC in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of

GCREC in cell lines is preferred. For example, sequences encoding GCREC can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding GCREC is inserted within a marker gene sequence, transformed cells containing sequences encoding GCREC can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GCREC under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding GCREC and that express GCREC may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or

chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of GCREC using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GCREC is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GCREC include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GCREC, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GCREC may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GCREC may be designed to contain signal sequences which direct secretion of GCREC through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities

(e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GCREC may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric GCREC protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of GCREC activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the GCREC encoding sequence and the heterologous protein sequence, so that GCREC may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled GCREC may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

GCREC of the present invention or fragments thereof may be used to screen for compounds that specifically bind to GCREC. At least one and up to a plurality of test compounds may be screened for specific binding to GCREC. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of GCREC, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which GCREC binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the

compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express GCREC, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing GCREC or cell membrane fractions which contain GCREC are then contacted with a test compound and binding, stimulation, or inhibition of activity of either GCREC or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with GCREC, either in solution or affixed to a solid support, and detecting the binding of GCREC to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

GCREC of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of GCREC. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for GCREC activity, wherein GCREC is combined with at least one test compound, and the activity of GCREC in the presence of a test compound is compared with the activity of GCREC in the absence of the test compound. A change in the activity of GCREC in the presence of the test compound is indicative of a compound that modulates the activity of GCREC. Alternatively, a test compound is combined with an in vitro or cell-free system comprising GCREC under conditions suitable for GCREC activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of GCREC may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding GCREC or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP

system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding GCREC may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding GCREC can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding GCREC is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress GCREC, e.g., by secreting GCREC in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of GCREC and G-protein coupled receptors. In addition, the expression of GCREC is closely associated with ovarian tumor, prostate, white blood cells, cerebellar, and brain tissues. Therefore, GCREC appears to play a role in cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections. In the treatment of disorders associated with increased GCREC expression or activity, it is desirable to decrease the expression or activity of GCREC. In the treatment of disorders associated with decreased GCREC expression or activity, it is desirable to increase the expression or activity of GCREC.

Therefore, in one embodiment, GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria,

polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer,

cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus.

In another embodiment, a vector capable of expressing GCREC or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified GCREC in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of GCREC may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of GCREC including, but not limited to, those listed above.

In a further embodiment, an antagonist of GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, cardiovascular, gastrointestinal, autoimmune/inflammatory, and metabolic disorders, and viral infections described above. In one aspect, an antibody which specifically binds GCREC may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express GCREC.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding GCREC may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of GCREC including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of GCREC may be produced using methods which are generally known in the art. In particular, purified GCREC may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GCREC. Antibodies to GCREC may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with GCREC or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to GCREC have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of GCREC amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to GCREC may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GCREC-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for GCREC may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

immunoassays typically involve the measurement of complex formation between GCREC and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering GCREC epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for GCREC. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of GCREC-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple GCREC epitopes, represents the average affinity, or avidity, of the antibodies for GCREC. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular GCREC epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the GCREC-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of GCREC, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of GCREC-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding GCREC, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding GCREC. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding GCREC. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense

sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding GCREC may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in GCREC expression or regulation causes disease, the expression of GCREC from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in GCREC are treated by constructing mammalian expression vectors encoding GCREC and introducing these vectors by mechanical means into GCREC-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic

gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of GCREC include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). GCREC may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding GCREC from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to GCREC expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding GCREC under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.

(1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding GCREC to cells which have one or more genetic abnormalities with respect to the expression of GCREC. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding GCREC to target cells which have one or more genetic abnormalities with respect to the expression of GCREC. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing GCREC to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev.

Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding GCREC to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for GCREC into the alphavirus genome in place of the capsid-coding region results in the production of a large number of GCREC-coding RNAs and the synthesis of high levels of GCREC in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of GCREC into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme

molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GCREC.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GCREC. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding GCREC. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased GCREC expression or activity, a compound which specifically inhibits expression of the

polynucleotide encoding GCREC may be therapeutically useful, and in the treament of disorders associated with decreased GCREC expression or activity, a compound which specifically promotes expression of the polynucleotide encoding GCREC may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding GCREC is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding GCREC are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding GCREC. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such

therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of GCREC, antibodies to GCREC, and mimetics, agonists, antagonists, or inhibitors of GCREC.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising GCREC or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, GCREC or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GCREC or fragments thereof, antibodies of GCREC, and agonists, antagonists or inhibitors of GCREC, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GCREC may be used for the diagnosis of disorders characterized by expression of GCREC, or in assays to monitor patients being treated with GCREC or agonists, antagonists, or inhibitors of GCREC. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for GCREC include methods which utilize the antibody and a label to detect GCREC in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A

wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring GCREC, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of GCREC expression. Normal or standard values for GCREC expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to GCREC under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of GCREC expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GCREC may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of GCREC may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of GCREC, and to monitor regulation of GCREC levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GCREC or closely related molecules may be used to identify nucleic acid sequences which encode GCREC. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding GCREC, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the GCREC encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:22-42 or from genomic sequences including promoters, enhancers, and introns of the GCREC gene.

Means for producing specific hybridization probes for DNAs encoding GCREC include the cloning of polynucleotide sequences encoding GCREC or GCREC derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GCREC may be used for the diagnosis of disorders associated with expression of GCREC. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart

disease, congenital heart disease, and complications of cardiac transplantation; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a metabolic disorder such as diabetes, obesity, and osteoporosis; and an infection by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, and tongavirus. The polynucleotide sequences encoding GCREC may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered GCREC expression. Such qualitative or quantitative

methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GCREC may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding GCREC may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding GCREC in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of GCREC, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding GCREC, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GCREC may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding

GCREC, or a fragment of a polynucleotide complementary to the polynucleotide encoding GCREC, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding GCREC are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of GCREC include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene

function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, GCREC, fragments of GCREC, or antibodies specific for GCREC may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested

compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of

at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for GCREC to quantify the levels of GCREC expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well

known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding GCREC may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding GCREC on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, GCREC, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between GCREC and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with GCREC, or fragments thereof, and washed. Bound GCREC is then detected by methods well known in the art. Purified GCREC can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GCREC specifically compete with a test compound for binding GCREC. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GCREC.

In additional embodiments, the nucleotide sequences which encode GCREC may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/186,854, U.S. Ser. No. 60/188,384, U.S. Ser. No. 60/190,453, U.S. Ser. No. 60/190,730 are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol

and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length

polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative G-protein coupled receptors were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode G-protein coupled receptors, the encoded polypeptides were analyzed by querying against PFAM models for G-protein coupled receptors. Potential G-protein coupled receptors were also identified by homology to Incyte cDNA sequences that had been annotated as G-protein coupled

receptors. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public

databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of GCREC Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:22-42 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:23 was mapped to chromosome 16 within the interval from 57.8 to 71.4 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene

and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding GCREC are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer,

cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding GCREC. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of GCREC Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon

membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse

transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522 , incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the GCREC-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring GCREC. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of GCREC. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the GCREC-encoding transcript.

XII. Expression of GCREC

Expression and purification of GCREC is achieved using bacterial or virus-based expression systems. For expression of GCREC in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express GCREC upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of GCREC in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding GCREC by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, GCREC is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from GCREC at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified GCREC obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, etc. where applicable.

XIII. Functional Assays

GCREC function is assessed by expressing the sequences encoding GCREC at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of GCREC on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding GCREC and either CD64 or CD64-GFP. CD64 and

CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding GCREC and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of GCREC Specific Antibodies

GCREC substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the GCREC amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-GCREC activity by, for example, binding the peptide or GCREC to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring GCREC Using Specific Antibodies

Naturally occurring or recombinant GCREC is substantially purified by immunoaffinity chromatography using antibodies specific for GCREC. An immunoaffinity column is constructed by covalently coupling anti-GCREC antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GCREC are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GCREC (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/GCREC binding (e.g.,'a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GCREC is collected.

XVI. Identification of Molecules Which Interact with GCREC

Molecules which interact with GCREC may include agonists and antagonists, as well as molecules involved in signal transduction, such as G proteins. GCREC, or a fragment thereof, is labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) A fragment of GCREC includes, for example, a fragment comprising one or more of the three extracellular loops, the extracellular N-terminal region, or the third intracellular loop. Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GCREC, washed, and any wells with labeled GCREC complex are assayed. Data obtained using different concentrations of GCREC are used to calculate values for the number, affinity, and association of GCREC with the candidate ligand molecules.

Alternatively, molecules interacting with GCREC are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech). GCREC may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential GCREC agonists or antagonists may be tested for activation or inhibition of GCREC receptor activity using the assays described in sections XVII and XVIII. Candidate molecules may be selected from known GPCR agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Methods for detecting interactions of GCREC with intracellular signal transduction molecules such as G proteins are based on the premise that internal segments or cytoplasmic domains from an orphan G protein-coupled seven transmembrane receptor may be exchanged with the analogous domains of a known G protein-coupled seven transmembrane receptor and used to identify the G-proteins and downstream signaling pathways activated by the orphan receptor domains (Kobilka, B.K. et al. (1988) Science 240:1310-1316). In an analogous fashion, domains of the orphan receptor may be cloned as a portion of a fusion protein and used in binding assays to demonstrate interactions with specific G proteins. Studies have shown that the third intracellular loop of G protein-coupled seven transmembrane receptors is important for G protein interaction and signal transduction (Conklin, B.R. et al. (1993) Cell 73:631-641). For example, the DNA fragment corresponding to the third intracellular loop of GCREC may be amplified by the polymerase chain reaction (PCR) and subcloned into a fusion vector such as pGEX (Pharmacia Biotech). The construct is transformed into an appropriate bacterial host, induced, and the fusion protein is purified from the cell lysate by glutathione-Sepharose 4B (Pharmacia Biotech) affinity chromatography.

For in vitro binding assays, cell extracts containing G proteins are prepared by extraction with 50 mM Tris, pH 7.8, 1 mM EGTA, 5 mM $MgCl_2$, 20 mM CHAPS, 20% glycerol, 10 μg of both aprotinin and leupeptin, and 20 μl of 50 mM phenylmethylsulfonyl fluoride. The lysate is incubated on ice for 45 min with constant stirring, centrifuged at 23,000 g for 15 min at 4°C, and the supernatant is collected. 750 μg of cell extract is incubated with glutathione S-transferase (GST) fusion protein beads for 2 h at 4°C. The GST beads are washed five times with phosphate-buffered saline. Bound G subunits are detected by [^{32}P]ADP-ribosylation with pertussis or cholera toxins. The reactions are terminated by the addition of SDS sample buffer (4.6% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (w/v) glycerol, 95.2 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromphenol blue). The [^{32}P]ADP-labeled proteins are separated on 10% SDS-PAGE gels, and autoradiographed. The separated proteins in these gels are transferred to nitrocellulose paper, blocked with blotto (5% nonfat dried milk, 50 mM Tris-HCl (pH 8.0), 2 mM $CaCl_2$, 80 mM NaCl, 0.02% NaN_3, and 0.2% Nonidet P-40) for 1 hour at room temperature, followed by incubation for 1.5 hours with Gα subtype selective antibodies (1:500; Calbiochem-Novabiochem). After three washes, blots are incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin (1:2000, Cappel, Westchester PA) and visualized by the chemiluminescence-based ECL method (Amersham Corp.).

XVII. Demonstration of GCREC Activity

An assay for GCREC activity measures the expression of GCREC on the cell surface. cDNA encoding GCREC is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using GCREC-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of GCREC expressed on the cell surface.

In the alternative, an assay for GCREC activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding GCREC is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor molecule. Varying amounts of GCREC ligand are then added to the cultured cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold GCREC ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the

concentration of GCREC producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for GCREC activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length GCREC is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing $1x10^5$ cells/well and incubated with inositol-free media and [^{3}H]myoinositol, 2 μCi/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of GCREC present in the transfected cells.

XVIII. Identification of GCREC Ligands

GCREC is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293 which have a good history of GPCR expression and which contain a wide range of G-proteins allowing for functional coupling of the expressed GCREC to downstream effectors. The transformed cells are assayed for activation of the expressed receptors in the presence of candidate ligands. Activity is measured by changes in intracellular second messengers, such as cyclic AMP or Ca^{2+}. These may be measured directly using standard methods well known in the art, or by the use of reporter gene assays in which a luminescent protein (e.g. firefly luciferase or green fluorescent protein) is under the transcriptional control of a promoter responsive to the stimulation of protein kinase C by the activated receptor (Milligan, G. et al. (1996) Trends Pharmacol. Sci. 17:235-237). Assay technologies are available for both of these second messenger systems to allow high throughput readout

in multi-well plate format, such as the adenylyl cyclase activation FlashPlate Assay (NEN Life Sciences Products), or fluorescent Ca^{2+} indicators such as Fluo-4 AM (Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In cases where the physiologically relevant second messenger pathway is not known, GCREC may be coexpressed with the G-proteins $G_{\alpha 15/16}$ which have been demonstrated to couple to a wide range of G-proteins (Offermanns, S. and M.I. Simon (1995) J. Biol. Chem. 270:15175-15180), in order to funnel the signal transduction of the GCREC through a pathway involving phospholipase C and Ca^{2+} mobilization. Alternatively, GCREC may be expressed in engineered yeast systems which lack endogenous GPCRs, thus providing the advantage of a null background for GCREC activation screening. These yeast systems substitute a human GPCR and G_{α} protein for the corresponding components of the endogenous yeast pheromone receptor pathway. Downstream signaling pathways are also modified so that the normal yeast response to the signal is converted to positive growth on selective media or to reporter gene expression (Broach, J.R. and J. Thorner (1996) Nature 384 (supp.):14-16). The receptors are screened against putative ligands including known GPCR ligands and other naturally occurring bioactive molecules. Biological extracts from tissues, biological fluids and cell supernatants are also screened.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
536482	1	536482CD1	22	536482CB1
1316020	2	1316020CD1	23	1316020CB1
2816437	3	2816437CD1	24	2816437CB1
2289894	4	2289894CD1	25	2289894CB1
7066050	5	7066050CD1	26	7066050CB1
5376785	6	5376785CD1	27	5376785CB1
3082743	7	3082743CD1	28	3082743CB1
7472361	8	7472361CD1	29	7472361CB1
7472363	9	7472363CD1	30	7472363CB1
7472364	10	7472364CD1	31	7472364CB1
7472434	11	7472434CD1	32	7472434CB1
7472435	12	7472435CD1	33	7472435CB1
7472438	13	7472438CD1	34	7472438CB1
7472439	14	7472439CD1	35	7472439CB1
7472440	15	7472440CD1	36	7472440CB1
7472443	16	7472443CD1	37	7472443CB1
7472445	17	7472445CD1	38	7472445CB1
7472446	18	7472446CD1	39	7472446CB1
7472451	19	7472451CD1	40	7472451CB1
7472456	20	7472456CD1	41	7472456CB1
7472457	21	7472457CD1	42	7472457CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
2	1316020	g927209	4.10E-21	alpha 1C adrenergic receptor isoform 2 [Homo sapiens] (Hirasawa,A., et al. (1995) Cloning, functional expression and tissue distribution of human alpha 1c-adrenoceptor splice variants. FEBS Lett. 363 (3), 256-260.)
4	2289894	g992582	3.20E-35	G protein-coupled seven-transmembrane receptor [Oryzias latipes] (Yasuoka,A., et al. (1995) Molecular cloning of a fish gene encoding a novel seven-transmembrane receptor related distantly to catecholamine, histamine, and serotonin receptors. Biochim. Biophys. Acta 1235, 467-469.)
5	7066050	g6531395	2.50E-175	growth factor-regulated G protein-coupled receptor Nrg-1 [Rattus norvegicus] (Glickman,M., et al. (1999) Molecular cloning, tissue-specific expression, and chromosomal localization of a novel nerve growth factor-regulated G-protein-coupled receptor, nrg-1. Mol. Cell. Neurosci. 14, 141-152.)
6	5376785	g460318	5.30E-72	G-protein coupled receptor [Mus musculus] (Harrigan,M.T. et al. (1991) Identification of a gene induced by glucocorticoids in murine T-cells: a potential G protein-coupled receptor. Mol. Endocrinol. 5: 1331-1338.)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
7	3082743	g6178006	5.50E-141	odorant receptor MOR83 [Mus musculus] (Tsuboi,A. et al. (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci. 19:8409-8418.)
8	7472361	g3927808	1.80E-85	olfactory receptor-like protein COR3'beta [Gallus gallus] (Reitman,M. et al. (1993) Primary sequence, evolution, and repetitive elements of the Gallus gallus (chicken) beta-globin cluster. Genomics 18: 616-626.)
9	7472363	g6532001	1.90E-90	odorant receptor S19 [Mus musculus]
10	7472364	g4680268	2.00E-95	odorant receptor S46 [Mus musculus] (Malnic,B. et al. (1999) Combinatorial receptor codes for odors. Cell 96: 713-723.)
11	7472434	g5869927	2.20E-98	olfactory receptor [Mus musculus]
12	7472435	g5869916	4.00E-138	olfactory receptor [Mus musculus]
13	7472438	g9963968	1.00E-141	odorant receptor M72 [Mus musculus] (Zheng,C. et al. (2000) Peripheral olfactory projections are differentially affected in mice deficient in a cyclic nucleotide-gated channel subunit. Neuron 26: 81-91.)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
14	7472439	g11692541	1.00E-123	odorant receptor K23 [Mus musculus] (Xie,S.Y. et al. (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. Mamm. Genome 11: 1070-1078.)
15	7472440	g11692535	1.00E-135	odorant receptor K21 [Mus musculus] (Xie,S.Y. et al. (2000) Characterization of a cluster comprising approximately 100 odorant receptor genes in mouse. Mamm. Genome 11: 1070-1078.)
16	7472443	g1246534	3.50E-82	olfactory receptor 4 [Gallus gallus] (Leibovici,M. et al. (1996) Avian olfactory receptors: differentiation of olfactory neurons under normal and experimental conditions. Dev. Biol. 175:118-131.)
17	7472445	g5453092	1.00E-102	olfactory receptor [Mus musculus domesticus] (Rouquier,S. et al. (2000) The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. U.S.A. 97: 2870-2874.)
18	7472446	g3983382	4.00E-65	olfactory receptor E3 [Mus musculus] (Krautwurst,D. et al. (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926.)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
19	7472451	g1256393	1.90E-97	taste bud receptor protein TB 641 [Rattus norvegicus] (Thomas,M.B. et al. (1996) Chemoreceptors expressed in taste, olfactory and male reproductive tissues. Gene 178: 1-5.)
20	7472456	g3983374	2.80E-82	olfactory receptor C6 [Mus musculus] (Krautwurst,D. et al. (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95: 917-926.)
21	7472457	g6178006	3.90E-92	odorant receptor MOR83 [Mus musculus] (Tsuboi,A. et al. (1999) Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. J. Neurosci. 19:8409-8418.)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	536482CD1	99	T34		Transmembrane domain: G50-L70	HMMER
					Bacterial rhodopsins signature: Y32-K83	PROFILESCAN
2	1316020CD1	139	S14 T61 S131		Signal peptide: M1-S57	SPSCAN
					G-protein coupled receptors family 2 signature: E54-G119	PROFILESCAN
					TA2R_HUMAN, BETA ISOFORM PD168643: S70-L135 (P-value = 4.3e-07)	BLAST-PRODOM
3	2816437CD1	82	T4 S58 S21 S75 S78	N46	Signal peptide: M1-S48	SPSCAN
					Melanocortin receptor family signature: L39-F56	BLIMPS-PRINTS
					Vasopressin V2 receptor signature: F56-S75	BLIMPS-PRINTS
4	2289894CD1	368	S305	N3 N8	Transmembrane domain: L18-P43	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G31-Y290	HMMER-PFAM
					G-protein coupled receptors signature BL00237: V81-P120; F181-Y192; L234-A260	BLIMPS-BLOCKS
					Thromboxane receptor signature PR00429: A199-L220	BLIMPS-PRINTS
					G-protein coupled receptors DM00013|P25962|31-357: A13-R201; P206-Y290	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7066050CD1	398	T79 T309 S340 S361 T22 T100 S146 S237 S363	N20	Transmembrane domains: V42-V60; V194-Y212	HMMER
					7 transmembrane receptor (rhodopsin family) domain: E53-Y306	HMMER-PFAM
					G-protein coupled receptors signature BL00237: L101-R140; P244-L270; N298-R314	BLIMPS-BLOCKS
					Rhodopsin-like GPCR superfamily signature PR00237: A38-G62; M71-N92; V115-M137 R150-L171; Y193-Y216; L249-L273 A288-R314	BLIMPS-PRINTS
					EDG1 orphan receptor signature PR00642: E13-R29; V60-L74; S96-V115 D274-F291	BLIMPS-PRINTS
					G-protein coupled receptors DM00013|P21453|39-326: L36-V321	BLAST-DOMO
					G-protein coupled receptor motif: A121-M137	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	5376785CD1	153	T20, S46, S110, S116, S144		G-PROTEIN COUPLED RECEPTORS; DM00013|P30731|65-361:I2-S91	BLAST-DOMO
					PROBABLE G PROTEINCOUPLED RECEPTOR FROM T-CELLS PRECURSOR GLUCOCORTICOIDINDUCED GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN SIGNAL ALTERNATIVE SPLICING: PD061453:C76-S153	BLAST-PRODOM
					Signal cleavage site: M1-C40	SPSCAN
					Transmembrane domain: T20-L38	HMMER
					7-transmembrane receptor (rhodopsin family): A12-Y75	HMMER-PFAM
					Visual pigments (opsins) retinal binding site: W62-R103	PROFILESCAN
					G-PROTEIN COUPLED RECEPTORS: BL00237C:R15-Y41; BL00237D:S67-R83	BLIMPS-BLOCKS
					Rhodopsin-like GPCR super family: PR00237A:M23-S47; PR00237E:K22-L45; PR00237F:T20-L44; PR00237G:Y57-R83	BLIMPS-PRINTS
					Probable G protein coupled receptor domain: PR01018I:C76-L89	BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	3082743CD1	313	S229 S67 T259 T163 S224 T288	N5 N65	Transmembrane domains: M29-A48; Q100-V118; Y193-A213	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-Y287	HMMER-PFAM
					G-protein coupled receptors signature BL00237: K90-P129; T279-K295	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-L148	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: L26-V50; M59-K80; L104-I126 A236-R260; K269-K295	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-K80; Y177-D191; L237-G252 V271-L282; T288-R302	BLIMPS-PRINTS
					OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: V246-Q305	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29710|15-301: F28-L301	BLAST-DOMO
					G-protein coupled receptors motif: A110-I126	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7472361CD1	315	S14 T31 S32 T101 S114 T155 S224 T63 T307 T354 Y351	N12 N61 N89	Transmembrane domains: F78-I96; I182-I210; I239-I262	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G88-I197; F252-Y338	HMMER-PFAM
					G-protein coupled receptors signature BL00237: H137-P176; F253-Y264; D278-S304 P330-R346	BLIMPS-BLOCKS
					Olfactory receptor signature PR00245: M106-R127; S224-N238; L284-I299	BLIMPS-PRINTS
					PUTATIVE GPROTEIN COUPLED RECEPTOR RA1C PD170483: V291-F353	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|G45774|18-309: V81-E347	BLAST-DOMO
					G-protein coupled receptors motif: M157-I173	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7472363CD1	356	T110 S190 T226 S232 S295	N5 N44	Transmembrane domains: Y37-S55; P60-L84; V204-I223	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G43-Y294	HMMER-PFAM
					G-protein coupled receptors signature BL00237: P92-P131; P286-R302	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F104-S154	PROFILESCAN
					Olfactory receptor signature PR00245: M61-T82; A179-D193; L240-V255	BLIMPS-PRINTS
					PUTATIVE GPROTEIN COUPLED RECEPTOR RA1C PD170483: I247-F309	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29707|18-306: E23-R302	BLAST-DOMO
					G-protein coupled receptors motif: L112-I128	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7472364CD1	311	T56 S69 T110 S179 T262 T309	N5 N44	Transmembrane domains: F33-I51; I137-I165; I194-I217	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G43-I152; F207-Y293	HMMER-PFAM
					G-protein coupled receptors signature BL00237: H92-P131; F208-Y219; D233-S259 P285-R301	BLIMPS-BLOCKS
					Olfactory receptor signature PR00245: M61-R82; S179-N193; L239-I254	BLIMPS-PRINTS
					PUTATIVE GPROTEIN COUPLED RECEPTOR RA1C PD170483: V246-F308	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|G45774|18-309: P20-E302	BLAST-DOMO
					G-protein coupled receptors motif: M112-I128	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7472434CD1	354	S340 S28 S86 S95 S3 T60 S64 T116 S316 S336	N93	Transmembrane domain: I218-A235	HMMER
					7 transmembrane receptor (rhodopsin family) domain: N93-Y315	HMMER-PFAM
					G-protein coupled receptors signature BL00237: R118-P157; T233-Y244; T307-K323	BLIMPS-BLOCKS
					G-protein coupled receptors signature: Y130-L175	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: T132-I154; L225-L248; K297-K323	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: N87-L108; Y203-D217; F264-G279 V299-L310; S316-L330	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L194-H270	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|P23275|17-306: N92-L330	BLAST-DOMO
					G-protein coupled receptors motif: T138-I154	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7472435CD1	319	T50 S68 T88 S277 S298	N5 N66	Signal peptide: M1-G42	SPSCAN
					Transmembrane domains: V30-L48; M60-L83; I198-T225	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G42-Y297	HMMER-PFAM
					G-protein coupled receptors signature BL00237: E91-P130; T289-K305	BLIMPS-BLOCKS
					G-protein coupled receptors signature: L104-A147	PROFILESCAN
					Olfactory receptor signature PR00245: M60-L81; F178-D192; F239-G254 I281-L292; S298-L312	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L167-L246	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|P23275|17-306: L18-L312	BLAST-DOMO
					G-protein coupled receptors motif: T111-I127	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7472438CD1	309	T8 S67 S156 S190 S228 T18 T78 S87 S290 T303	N5 N65	Transmembrane domains: F28-L48; M98-D121	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-Y289	HMMER-PFAM
					G-protein coupled receptors signature BL00237: N90-P129; I281-K297	BLIMPS-BLOCKS
					G-protein coupled receptors signature: Y102-A150	PROFILESCAN
					Olfactory receptor signature PR00245: M59-K80; Y176-S190; F237-G252 A273-L284; S290-L304	BLIMPS-PRINTS
					OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L244	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S51356|18-307: L17-V300	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7472439CD1	310	S7 S66 S228 T77 S86 S290	N5 N64 N164 N189 N227	Transmembrane domains: L26-I45; M58-T77; P128-L145 E195-I220	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G40-Y289	HMMER-PFAM
					G-protein coupled receptors signature BL00237: N89-P128; V281-K297	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F101-G151	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: P25-G49; M58-K79; F103-I125 V198-L221; K271-K297	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M58-K79; F176-S190; F237-G252 S273-L284; S290-L304	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L165-I245	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29709|11-299: T17-G305	BLAST-DOMO
					G-protein coupled receptors motif: S109-I125	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7472440CD1	311	S91 S67 T78 S291	N5 N65	Transmembrane domains: F29-T47; Q100-M118; P129-L146	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-Y290	HMMER-PFAM
					G-protein coupled receptors signature BL00237: N90-P129; V282-K298	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-G150	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: P26-R50; M59-K80; F104-I126 V199-L222; K272-K298	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-K80; Y177-S191; F238-G253 S274-L285; S291-L305	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: V166-I246	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29709|11-299: T18-L305	BLAST-DOMO
					G-protein coupled receptors motif: S110-I126	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7472443CD1	314	S67 S267 T87 S264 S291	N5 N210	Transmembrane domains: L29-Y56; F101-D121; I197-I221 H244-S262	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-Y290	HMMER-PFAM
					G-protein coupled receptors signature BL00237: T90-P129; I282-K298	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-V150	PROFILESCAN
					Olfactory receptor signature PR00245: M59-Q80; F177-D191; F238-G253 V274-L285; S291-A305	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-I245	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S51356|18-307: P21-A300	BLAST-DOMO
					G-protein coupled receptors motif: T110-I126	MOTIFS
					ATP/GTP binding site (P-loop) A83-T90	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7472445CD1	346	T123 S32 S99 S188 T298 T25 T169 S177 S323 S343	N37 N97	Transmembrane domains: I61-L83; I133-D153; M229-L247	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G73-Y322	HMMER-PFAM
					G-protein coupled receptors signature BL00237: K122-P161; T314-K330	BLIMPS-BLOCKS
					G-protein coupled receptors signature: I134-F178	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: L58-F82; M91-Q112; F136-V158 I231-I254; K304-K330	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M91-Q112; Y209-D223; F270-G285; I306-L317; S323-V337	BLIMPS-PRINTS
					OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: V279-K340	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|P23275|17-306: Q56-G338	BLAST-DOMO
					G-protein coupled receptors motif: T142-V158	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7472446CD1	316	S67 S193 T268 T78 S137 T192 S267 S291	N5 N19	Transmembrane domains: L34-A53; E196-Y218	HMMER
					7 transmembrane receptor (rhodopsin family) domain: S41-Y290	HMMER-PFAM
					G-protein coupled receptors signature BL00237: N90-P129; T282-M298	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-T147	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: L26-T50; M59-K80; A104-I126 A140-V161; V199-L222; A237-L261 N272-M298	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-K80; L177-D191; L238-G253 I274-L285; S291-L305	BLIMPS-PRINTS
					OLFACTORY RECEPTOR PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD149621: T246-K307	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|P23275|17-306: A29-G306	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7472451CD1	453	S81 T213 T331 T151	N19 N441	Signal peptide: M1-G55	SPSCAN
					Transmembrane domains: L43-I60; I217-V236	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G55-I263	HMMER-PFAM
					G-protein coupled receptors signature BL00237: R104-P143; T294-K310	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F116-T162	PROFILESCAN
					Olfactory receptor signature PR00245: M73-K94; I191-D205; F252-V267 V286-L297; T303-G317	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L180-L259	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29710|15-301: L41-L316	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7472456CD1	323	S49 S67 S188 T230 T40 T291 S320	N5 N65	Transmembrane domains: F25-I45; I57-I85; Y102-D121 L194-Y218	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-F290	HMMER-PFAM
					G-protein coupled receptors signature BL00237: H90-P129; T282-Q298	BLIMPS-BLOCKS
					G-protein coupled receptors signature: Y102-A147	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: M59-K80; Y104-I126; A199-L222 K272-Q298	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-K80; F177-D191; F238-G253 A274-L285; T291-L305	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-L245	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|P23267|20-309: F17-L305	BLAST-DOMO
					G-protein coupled receptors motif: T110-I126	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphory-lation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7472457CD1	318	S67 T288	N5 N65 N137	Transmembrane domain: V30-V49	HMMER
					7 transmembrane receptor (rhodopsin family) domain: G41-Y287	HMMER-PFAM
					G-protein coupled receptors signature BL00237: P90-P129; T279-I295	BLIMPS-BLOCKS
					G-protein coupled receptors signature: F102-A147	PROFILESCAN
					Rhodopsin-like GPCR superfamily signature PR00237: V26-T50; M59-R80; F104-I126 V140-A161; M199-L222; A236-R260 K269-I295	BLIMPS-PRINTS
					Olfactory receptor signature PR00245: M59-R80; F177-D191; L237-V252 V271-L282; T288-W302	BLIMPS-PRINTS
					RECEPTOR OLFACTORY PROTEIN RECEPTORLIKE GPROTEIN COUPLED TRANSMEMBRANE GLYCOPROTEIN MULTIGENE FAMILY PD000921: L166-I244	BLAST-PRODOM
					G-PROTEIN COUPLED RECEPTORS DM00013|S29710|15-301: L17-W302	BLAST-DOMO

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	536482CB1	1348	1-174, 812-868	6871486H1 (BRAGNON02)	218	950
				1347882T6 (PROSNOT11)	457	1058
				536482R6 (LNODNOT02)	1	436
				7087611H1 (BRAUTDR03)	798	1348
23	1316020CB1	1446	1-750, 851-930, 771-832	5599305F8 (UTRENON03)	716	1330
				1836450R6 (BRAINON01)	1174	1446
				6442433H1 (BRAENOT02)	833	1440
				7121816H1 (BRAHNOE01)	1	445
				5174595H1 (EPIBTXT01)	325	619
				5575376H1 (BRAPNOT04)	508	769
24	2816437CB1	1463	155-231, 1427-1463, 644-800	2780847F6 (OVARTUT03)	250	716
				2816437H1 (BRSTNOT14)	1	276
				70171099V1	664	1054
				6915195H1 (PITUDIR01)	880	1463
				70173319V1	723	1225
				2943542H2 (BRAITUT23)	585	722
25	2289894CB1	1435	477-681, 1289-1435	FL2289894_00001	63	1435
				g5743982	1	455
26	7066050CB1	2147	629-1512, 1-141	7066050H1 (BRATNOR01)	1	512
				FL7066050_00001	11	2132
				70700621V1	1617	2147
27	5376785CB1	1989	1493-1989, 1-53, 875-990, 157-485, 1338-1432, 1095-1240	7101536H1 (BRAWTDR02)	712	1256
				71217673V1	1242	1989
				70822278V1	1	602
				70818975V1	985	1467
				71217707V1	485	1122
28	3082743CB1	942	897-942	GBI.g2121229.raw	1	360
				FL3082743_00001	121	942
29	7472361CB1	948	577-948	GNN.g6671985_006	1	948
30	7472363CB1	1071	554-1071	GNN.g6671985_018	1	1071

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7472364CB1	1001	1-55, 435-1001	GNN.g6671985_022	1	1001
32	7472434CB1	1065	1-356, 1002-1065, 561-594	GNN.g6630753_006	1	1065
33	7472435CB1	963	1-27, 762-963	GNN.g6630753_008	1	963
34	7472438CB1	1101	1-138, 582-712	GNN.g6635275_002	1	1101
35	7472439CB1	933		GNN.g6635275_006	1	933
36	7472440CB1	936	425-544	GNN.g6635275_018	1	936
37	7472443CB1	945	922-945	GNN.g6642708_020	1	945
38	7472445CB1	1041	1008-1041, 81-107	GNN.g6648400_008	1	1041
39	7472446CB1	951	920-951, 578-607	GNN.g6648400_012	1	951
40	7472451CB1	1395	1-87, 1208-1395, 857-1059	GNN.g6648431_006	1	1395
41	7472456CB1	972	462-972	GNN.g6648431_016	1	972
42	7472457CB1	957	916-957	GNN.g6648431_018	1	957

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
22	536482CB1	PROSNOT11
23	1316020CB1	BRAINOY02
24	2816437CB1	OVARTUT03
25	2289894CB1	BRAINON01
26	7066050CB1	LEUKNOT02
27	5376785CB1	BRAXNOT01

Table 6

Library	Vector	Library Description
BRAINON01	PSPORT1	Library was constructed and normalized from 4.88 million independent clones from RNA which was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRAINOY02	pINCY	This large size-fractionated and normalized library was constructed using pooled cDNA generated using mRNA isolated from midbrain, inferior temporal cortex, medulla, and posterior parietal cortex tissues removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. Scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver. 2.8 x 10e5 independent clones from this size-selected library were normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAXNOT01	pINCY	Library was constructed using RNA isolated from cerebellar tissue removed from a 70-year-old male. Patient history included chronic obstructive airways disease and left ventricular failure.
LEUKNOT02	pINCY	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).

Table 6 (cont.)

Library	Vector	Library Description
OVARTUT03	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 52-year-old mixed ethnicity female during a total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal and lymphatic structure biopsy, regional lymph node excision, and peritoneal tissue destruction. Pathology indicated an invasive grade 3 (of 4) seroanaplastic carcinoma forming a mass in the left ovary. Multiple tumor implants were present on the surface of the left ovary and fallopian tube, right ovary and fallopian tube, posterior surface of the uterus, and cul-de-sac. The endometrium was atrophic. Multiple (2) leiomyomata were identified, one subserosal and 1 intramural. Pathology also indicated a metastatic grade 3 seroanaplastic carcinoma involving the omentum, cul-de-sac peritoneum, left broad ligament peritoneum, and mesentery colon. Patient history included breast cancer, chronic peptic ulcer, and joint pain. Family history included colon cancer, cerebrovascular disease, breast cancer, type II diabetes, esophagus cancer, and depressive disorder.
PROSNOT11	pINCY	Library was constructed using RNA isolated from the prostate tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:22-42.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

18. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional GCREC, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
LAL, Preeti
TANG, Y. Tom
PATTERSON, Chandra
YAO, Monique G.
SHIH, Leo L.
TRIBOULEY, Catherine
LU, Dyung Aina M.
YUE, Henry
KHAN, Farrah A.
POLICKY, Jennifer L.
AU-YOUNG, Janice
YANG, Junming
HARLAND, Lee
WALSH, Roderick T.
LO, Terence P.
BOROWSKY, Mark L.

<120> G-PROTEIN COUPLED RECEPTORS

<130> PI-0044 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/186,854; 60/188,384; 60/190,453; 60/190,730
<151> 2000-03-03; 2000-03-10; 2000-03-17; 2000-03-20

<160> 42
<170> PERL Program

<210> 1
<211> 99
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 536482CD1

<400> 1
Met Ala Glu Gly Gly Phe Asp Pro Cys Glu Cys Val Cys Ser His
1 5 10 15
Glu His Ala Met Arg Arg Leu Ile Asn Leu Leu Arg Gln Ser Gln
20 25 30
Ser Tyr Cys Thr Asp Thr Glu Cys Leu Gln Glu Leu Pro Gly Pro
35 40 45
Ser Gly Asp Asn Gly Ile Ser Val Thr Met Ile Leu Val Ala Trp
50 55 60
Met Val Ile Ala Leu Ile Leu Phe Leu Leu Arg Pro Pro Asn Leu
65 70 75
Arg Gly Ser Ser Leu Pro Gly Lys Pro Thr Ser Pro His Asn Gly
80 85 90
Gln Asp Pro Pro Ala Pro Pro Val Asp
95

<210> 2
<211> 139
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1316020CD1

<400> 2
Met Gly His Pro Arg Ala Ile Gln Pro Ser Val Phe Phe Ser Pro

```
  1               5                  10                  15
Tyr Asp Val His Phe Leu Leu Tyr Pro Ile Arg Cys Pro Tyr Leu
                20                  25                  30
Lys Ile Gly Arg Phe His Ile Lys Leu Lys Gly Leu His Phe Leu
                35                  40                  45
Phe Ser Phe Leu Phe Phe Phe Phe Glu Thr Gln Ser His Ser Val
                50                  55                  60
Thr Arg Leu Glu Cys Ser Gly Thr Ile Ser Ala His Cys Asn Leu
                65                  70                  75
Cys Leu Pro Gly Ser Ser Asn Ser Pro Ala Ser Ala Ser Gln Val
                80                  85                  90
Ala Gly Thr Thr Gly Thr Cys His His Ala Gln Leu Ile Phe Val
                95                 100                 105
Phe Leu Ala Glu Met Gly Phe His His Ile Gly Gln Asp Gly Leu
               110                 115                 120
Asp Leu Asn Leu Val Ile His Pro Pro Arg Ser Pro Lys Ala Leu
               125                 130                 135
Gly Leu Gln Ala
```

<210> 3
<211> 82
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2816437CD1

<400> 3

```
Met Gly Lys Thr Pro Ser Glu Ala Gln Asp Ser Leu Val Thr Phe
  1               5                  10                  15
Gln Phe Ala Asp Thr Ser Val Lys Val Ser Trp Glu Thr Ser Ala
                20                  25                  30
Leu Gly Ser Ser Ser Val Val Leu Leu Thr Leu Pro Val Lys Gln
                35                  40                  45
Asn Leu Ser Ser Val Cys Ile Gly Phe His Phe Leu Ser Pro Pro
                50                  55                  60
Glu Glu Trp Lys Ala Thr Ala Gln Ser Leu Leu Met Phe Trp Ser
                65                  70                  75
Glu Arg Ser Leu Lys Val Met
                80
```

<210> 4
<211> 368
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2289894CD1

<400> 4

```
Met Ala Asn Ser Thr Gly Leu Asn Ala Ser Glu Val Ala Gly Ser
  1               5                  10                  15
Leu Gly Leu Ile Leu Ala Ala Val Val Glu Val Gly Ala Leu Leu
                20                  25                  30
Gly Asn Gly Ala Leu Leu Val Val Val Leu Arg Thr Pro Gly Leu
                35                  40                  45
Arg Asp Ala Leu Tyr Leu Ala His Leu Cys Val Val Asp Leu Leu
                50                  55                  60
Ala Ala Ala Ser Ile Met Pro Leu Gly Leu Leu Ala Ala Pro Pro
                65                  70                  75
Pro Gly Leu Gly Arg Val Arg Leu Gly Pro Ala Pro Cys Arg Ala
                80                  85                  90
Ala Arg Phe Leu Ser Ala Ala Leu Leu Pro Ala Cys Thr Leu Gly
                95                 100                 105
```

```
Val Ala Ala Leu Gly Leu Ala Arg Tyr Arg Leu Ile Val His Pro
                110                 115                 120
Leu Arg Pro Gly Ser Arg Pro Pro Pro Val Leu Val Leu Thr Ala
                125                 130                 135
Val Trp Ala Ala Ala Gly Leu Leu Gly Ala Leu Ser Leu Leu Gly
                140                 145                 150
Pro Pro Pro Ala Pro Pro Pro Ala Pro Ala Arg Cys Ser Val Leu
                155                 160                 165
Ala Gly Gly Leu Gly Pro Phe Arg Pro Leu Trp Ala Leu Leu Ala
                170                 175                 180
Phe Ala Leu Pro Ala Leu Leu Leu Leu Gly Ala Tyr Gly Gly Ile
                185                 190                 195
Phe Val Val Ala Arg Arg Ala Ala Leu Arg Pro Pro Arg Pro Ala
                200                 205                 210
Arg Gly Ser Arg Leu Arg Ser Asp Ser Leu Asp Ser Arg Leu Ser
                215                 220                 225
Ile Leu Pro Pro Leu Arg Pro Arg Leu Pro Gly Gly Lys Ala Ala
                230                 235                 240
Leu Ala Pro Ala Leu Ala Val Gly Gln Phe Ala Ala Cys Trp Leu
                245                 250                 255
Pro Tyr Gly Cys Ala Cys Leu Ala Pro Ala Ala Arg Ala Ala Glu
                260                 265                 270
Ala Glu Ala Ala Val Thr Trp Val Ala Tyr Ser Ala Phe Ala Ala
                275                 280                 285
His Pro Phe Leu Tyr Gly Leu Leu Gln Arg Pro Val Arg Leu Ala
                290                 295                 300
Leu Gly Arg Leu Ser Arg Arg Ala Leu Pro Gly Pro Val Arg Ala
                305                 310                 315
Cys Thr Pro Gln Ala Trp His Pro Arg Ala Leu Leu Gln Cys Leu
                320                 325                 330
Gln Arg Pro Pro Glu Gly Pro Ala Val Gly Pro Ser Glu Ala Pro
                335                 340                 345
Glu Gln Thr Pro Glu Leu Ala Gly Gly Arg Ser Pro Ala Tyr Gln
                350                 355                 360
Gly Pro Pro Glu Ser Ser Leu Ser
                365
```

<210> 5
<211> 398
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7066050CD1

<400> 5

```
Met Glu Ser Gly Leu Leu Arg Pro Ala Pro Val Ser Glu Val Ile
  1               5                  10                  15
Val Leu His Tyr Asn Tyr Thr Gly Lys Leu Arg Gly Ala Arg Tyr
                20                  25                  30
Gln Pro Gly Ala Gly Leu Arg Ala Asp Ala Val Val Cys Leu Ala
                35                  40                  45
Val Cys Ala Phe Ile Val Leu Glu Asn Leu Ala Val Leu Leu Val
                50                  55                  60
Leu Gly Arg His Pro Arg Phe His Ala Pro Met Phe Leu Leu Leu
                65                  70                  75
Gly Ser Leu Thr Leu Ser Asp Leu Leu Ala Gly Ala Ala Tyr Ala
                80                  85                  90
Ala Asn Ile Leu Leu Ser Gly Pro Leu Thr Leu Lys Leu Ser Pro
                95                 100                 105
Ala Leu Trp Phe Ala Arg Glu Gly Gly Val Phe Val Ala Leu Thr
                110                 115                 120
Ala Ser Val Leu Ser Leu Leu Ala Ile Ala Leu Glu Arg Ser Leu
                125                 130                 135
Thr Met Ala Arg Arg Gly Pro Ala Pro Val Ser Ser Arg Gly Arg
                140                 145                 150
```

```
Thr Leu Ala Met Ala Ala Ala Ala Trp Gly Val Ser Leu Leu Leu
                155                 160                 165
Gly Leu Leu Pro Ala Leu Gly Trp Asn Cys Leu Gly Arg Leu Asp
            170                 175                 180
Ala Cys Ser Thr Val Leu Pro Leu Tyr Ala Lys Ala Tyr Val Leu
        185                 190                 195
Phe Cys Val Leu Ala Phe Val Gly Ile Leu Ala Ala Ile Cys Ala
    200                 205                 210
Leu Tyr Ala Arg Ile Tyr Cys Gln Val Arg Ala Asn Ala Arg Arg
215                 220                 225
Leu Pro Ala Arg Pro Gly Thr Ala Gly Thr Thr Ser Thr Arg Ala
                230                 235                 240
Arg Arg Lys Pro Arg Ser Leu Ala Leu Leu Arg Thr Leu Ser Val
            245                 250                 255
Val Leu Leu Ala Phe Val Ala Cys Trp Gly Pro Leu Phe Leu Leu
        260                 265                 270
Leu Leu Leu Asp Val Ala Cys Pro Ala Arg Thr Cys Pro Val Leu
    275                 280                 285
Leu Gln Ala Asp Pro Phe Leu Gly Leu Ala Met Ala Asn Ser Leu
290                 295                 300
Leu Asn Pro Ile Ile Tyr Thr Leu Thr Asn Arg Asp Leu Arg His
305                 310                 315
Ala Leu Leu Arg Leu Val Cys Cys Gly Arg His Ser Cys Gly Arg
                320                 325                 330
Asp Pro Ser Gly Ser Gln Gln Ser Ala Ser Ala Ala Glu Ala Ser
            335                 340                 345
Gly Gly Leu Arg Arg Cys Leu Pro Pro Gly Leu Asp Gly Ser Phe
        350                 355                 360
Ser Gly Ser Glu Arg Ser Ser Pro Gln Arg Asp Gly Leu Asp Thr
    365                 370                 375
Ser Gly Ser Thr Gly Ser Pro Gly Ala Pro Thr Ala Ala Arg Thr
380                 385                 390
Leu Val Ser Glu Pro Ala Ala Asp
395
```

<210> 6
<211> 153
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5376785CD1

<400> 6

```
Met Ile Gly Asp Val Thr Thr Glu Gln Tyr Phe Ala Leu Arg Arg
  1               5                  10                  15
Lys Lys Lys Lys Thr Ile Lys Met Leu Met Leu Val Val Val Leu
            20                  25                  30
Phe Ala Leu Cys Trp Phe Pro Leu Asn Cys Tyr Val Leu Leu Leu
        35                  40                  45
Ser Ser Lys Val Ile Arg Thr Asn Asn Ala Leu Tyr Phe Ala Phe
    50                  55                  60
His Trp Phe Ala Met Ser Ser Thr Cys Tyr Asn Pro Phe Ile Tyr
65                  70                  75
Cys Trp Leu Asn Glu Asn Phe Arg Ile Glu Leu Lys Ala Leu Leu
                80                  85                  90
Ser Met Cys Gln Arg Pro Pro Lys Pro Gln Glu Asp Arg Gln Pro
            95                  100                 105
Ser Pro Val Pro Ser Phe Arg Val Ala Trp Thr Glu Lys Asn Asp
        110                 115                 120
Gly Gln Arg Ala Pro Leu Ala Asn Asn Leu Leu Pro Thr Ser Gln
    125                 130                 135
Leu Gln Ser Gly Lys Thr Asp Leu Ser Ser Val Glu Pro Ile Val
140                 145                 150
Thr Met Ser
```

<210> 7
<211> 313
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3082743CD1

<400> 7
Met Asp Ser Leu Asn Gln Thr Arg Val Thr Glu Phe Val Phe Leu
1 5 10 15
Gly Leu Thr Asp Asn Arg Val Leu Glu Met Leu Phe Phe Met Ala
20 25 30
Phe Ser Ala Ile Tyr Met Leu Thr Leu Ser Gly Asn Ile Leu Ile
35 40 45
Ile Ile Ala Thr Val Phe Thr Pro Ser Leu His Thr Pro Met Tyr
50 55 60
Phe Phe Leu Ser Asn Leu Ser Phe Ile Asp Ile Cys His Ser Ser
65 70 75
Val Thr Val Pro Lys Met Leu Glu Gly Leu Leu Leu Glu Arg Lys
80 85 90
Thr Ile Ser Phe Asp Asn Cys Ile Thr Gln Leu Phe Phe Leu His
95 100 105
Leu Phe Ala Cys Ala Glu Ile Phe Leu Leu Ile Ile Val Ala Tyr
110 115 120
Asp Arg Tyr Val Ala Ile Cys Thr Pro Leu His Tyr Pro Asn Val
125 130 135
Met Asn Met Arg Val Cys Ile Gln Leu Val Phe Ala Leu Trp Leu
140 145 150
Gly Gly Thr Val His Ser Leu Gly Gln Thr Phe Leu Thr Ile Arg
155 160 165
Leu Pro Tyr Cys Gly Pro Asn Ile Ile Asp Ser Tyr Phe Cys Asp
170 175 180
Val Pro Leu Val Ile Lys Leu Ala Cys Thr Asp Thr Tyr Leu Thr
185 190 195
Gly Ile Leu Ile Val Thr Asn Ser Gly Thr Ile Ser Leu Ser Cys
200 205 210
Phe Leu Ala Val Val Thr Ser Tyr Met Val Ile Leu Val Ser Leu
215 220 225
Arg Lys His Ser Ala Glu Gly Arg Gln Lys Ala Leu Ser Thr Cys
230 235 240
Ser Ala His Phe Met Val Val Ala Leu Phe Phe Gly Pro Cys Ile
245 250 255
Phe Ile Tyr Thr Arg Pro Asp Thr Ser Phe Ser Ile Asp Lys Val
260 265 270
Val Ser Val Phe Tyr Thr Val Val Thr Pro Leu Leu Asn Pro Phe
275 280 285
Ile Tyr Thr Leu Arg Asn Glu Glu Val Lys Ser Ala Met Lys Gln
290 295 300
Leu Arg Gln Arg Gln Val Phe Phe Thr Lys Ser Tyr Thr
305 310

<210> 8
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472361CD1

<400> 8
Met Gly Asp Trp Asn Asn Ser Asp Ala Val Glu Pro Ile Phe Ile
1 5 10 15
Leu Arg Gly Phe Pro Gly Leu Glu Tyr Val His Ser Trp Leu Ser
20 25 30

```
Ile Leu Phe Cys Leu Ala Tyr Leu Val Ala Phe Met Gly Asn Val
                35                  40                  45
Thr Ile Leu Ser Val Ile Trp Ile Glu Ser Ser Leu His Gln Pro
                50                  55                  60
Met Tyr Tyr Phe Ile Ser Ile Leu Ala Val Asn Asp Leu Gly Met
                65                  70                  75
Ser Leu Ser Thr Leu Pro Thr Met Leu Ala Val Leu Trp Leu Asp
                80                  85                  90
Ala Pro Glu Ile Gln Ala Ser Ala Cys Tyr Ala Gln Leu Phe Phe
                95                  100                 105
Ile His Thr Phe Thr Phe Leu Glu Ser Ser Val Leu Leu Ala Met
                110                 115                 120
Ala Phe Asp Arg Phe Val Ala Ile Cys His Pro Leu His Tyr Pro
                125                 130                 135
Thr Ile Leu Thr Asn Ser Val Ile Gly Lys Ile Gly Leu Ala Cys
                140                 145                 150
Leu Leu Arg Ser Leu Gly Val Val Leu Pro Thr Pro Leu Leu Leu
                155                 160                 165
Arg His Tyr His Tyr Cys His Gly Asn Ala Leu Ser His Ala Phe
                170                 175                 180
Cys Leu His Gln Asp Val Leu Arg Leu Ser Cys Thr Asp Ala Arg
                185                 190                 195
Thr Asn Ser Ile Tyr Gly Leu Cys Val Val Ile Ala Thr Leu Gly
                200                 205                 210
Val Asp Ser Ile Phe Ile Leu Leu Ser Tyr Val Leu Ile Leu Asn
                215                 220                 225
Thr Val Leu Asp Ile Ala Ser Arg Glu Glu Gln Leu Lys Ala Leu
                230                 235                 240
Asn Thr Cys Val Ser His Ile Cys Val Val Leu Ile Phe Phe Val
                245                 250                 255
Pro Val Ile Gly Val Ser Met Val His Arg Phe Gly Lys His Leu
                260                 265                 270
Ser Pro Ile Val His Ile Leu Met Ala Asp Ile Tyr Leu Leu Leu
                275                 280                 285
Pro Pro Val Leu Asn Pro Ile Val Tyr Ser Val Arg Thr Lys Gln
                290                 295                 300
Ile Arg Leu Gly Ile Leu His Lys Phe Val Leu Arg Arg Arg Phe
                305                 310                 315
```

<210> 9
<211> 356
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472363CD1

<400> 9

```
Met Ile His Gly Gly Asp Pro Asn Ile Asn Ile Asn Arg Ser Leu
  1               5                  10                  15
Glu Glu Ala His Ser Asn Leu Met Asp Asn Val Glu Gly Phe Lys
                20                  25                  30
Thr Ser Val Glu Glu Ala Ala Ala Asp Met Val Glu Ile Ala Arg
                35                  40                  45
Glu Met Glu Leu Glu Val Lys Pro Glu Asp Gly Thr Glu Cys Cys
                50                  55                  60
Asn Leu Thr Thr Lys Gly Leu Glu Asp Phe His Met Trp Ile Ser
                65                  70                  75
Gly Pro Phe Cys Ser Val Tyr Leu Val Ala Leu Leu Gly Asn Ala
                80                  85                  90
Thr Ile Leu Leu Val Ile Lys Val Glu Gln Thr Leu Arg Glu Pro
                95                  100                 105
Met Phe Tyr Phe Leu Ala Ile Leu Ser Thr Ile Asp Leu Ala Leu
                110                 115                 120
Ser Ala Thr Ser Val Pro Arg Met Leu Gly Ile Phe Trp Phe Asp
```

```
                125                 130                 135
Ala His Glu Ile Asn Tyr Gly Ala Cys Val Ala Gln Met Phe Leu
                140                 145                 150
Ile His Ala Phe Thr Gly Met Glu Ala Glu Val Leu Leu Ala Met
                155                 160                 165
Ala Phe Asp Arg Tyr Val Ala Ile Cys Ala Pro Leu His Tyr Ala
                170                 175                 180
Thr Ile Leu Thr Ser Leu Val Leu Val Gly Ile Ser Met Cys Ile
                185                 190                 195
Val Ile Arg Pro Val Leu Leu Thr Leu Pro Met Val Tyr Leu Ile
                200                 205                 210
Tyr Arg Leu Pro Phe Cys Gln Ala His Ile Ile Ala His Ser Tyr
                215                 220                 225
Cys Glu His Met Gly Ile Ala Lys Leu Ser Cys Gly Asn Ile Arg
                230                 235                 240
Ile Asn Gly Ile Tyr Gly Leu Phe Val Val Ser Phe Phe Val Leu
                245                 250                 255
Asn Leu Val Leu Ile Gly Ile Ser Tyr Val Tyr Ile Leu Arg Ala
                260                 265                 270
Val Phe Arg Leu Pro Ser His Asp Ala Gln Leu Lys Ala Leu Ser
                275                 280                 285
Thr Cys Gly Ala His Val Gly Val Ile Cys Val Phe Tyr Ile Pro
                290                 295                 300
Ser Val Phe Ser Phe Leu Thr His Arg Phe Gly His Gln Ile Pro
                305                 310                 315
Gly Tyr Ile His Ile Leu Val Ala Asn Leu Tyr Leu Ile Ile Pro
                320                 325                 330
Pro Ser Leu Asn Pro Ile Ile Tyr Gly Val Arg Thr Lys Gln Ile
                335                 340                 345
Arg Glu Arg Val Leu Tyr Val Phe Thr Lys Lys
                350                 355
```

<210> 10
<211> 311
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472364CD1

<400> 10

```
Met Phe Tyr His Asn Lys Ser Ile Phe His Pro Val Thr Phe Phe
  1                 5                  10                  15
Leu Ile Gly Ile Pro Gly Leu Glu Asp Phe His Met Trp Ile Ser
                 20                  25                  30
Gly Pro Phe Cys Ser Val Tyr Leu Val Ala Leu Leu Gly Asn Ala
                 35                  40                  45
Thr Ile Leu Leu Val Ile Lys Val Glu Gln Thr Leu Arg Glu Pro
                 50                  55                  60
Met Phe Tyr Phe Leu Ala Ile Leu Ser Thr Ile Asp Leu Ala Leu
                 65                  70                  75
Ser Ala Thr Ser Val Pro Arg Met Leu Gly Ile Phe Trp Phe Asp
                 80                  85                  90
Ala His Glu Ile Asn Tyr Gly Ala Cys Val Ala Gln Met Phe Leu
                 95                 100                 105
Ile His Ala Phe Thr Gly Met Glu Ala Glu Val Leu Leu Ala Met
                110                 115                 120
Ala Phe Asp Arg Tyr Val Ala Ile Cys Ala Pro Leu His Tyr Ala
                125                 130                 135
Thr Ile Leu Thr Ser Leu Val Leu Val Gly Ile Ser Met Cys Ile
                140                 145                 150
Val Ile Arg Pro Val Leu Leu Thr Leu Pro Met Val Tyr Leu Ile
                155                 160                 165
Tyr Arg Leu Pro Phe Cys Gln Ala His Ile Ile Ala His Ser Tyr
                170                 175                 180
Cys Glu His Met Gly Ile Ala Lys Leu Ser Cys Gly Asn Ile Arg
```

```
                185                 190                 195
Ile Asn Gly Ile Tyr Gly Leu Phe Val Val Ser Phe Phe Val Leu
                200                 205                 210
Asn Leu Val Leu Ile Gly Ile Ser Tyr Val Tyr Ile Leu Arg Ala
                215                 220                 225
Val Phe Arg Leu Pro Ser His Asp Ala Gln Leu Lys Ala Leu Ser
                230                 235                 240
Thr Cys Gly Ala His Val Gly Val Ile Cys Val Phe Tyr Ile Pro
                245                 250                 255
Ser Val Phe Ser Phe Leu Thr His Arg Phe Gly His Gln Ile Pro
                260                 265                 270
Gly Tyr Ile His Ile Leu Val Ala Asn Leu Tyr Leu Ile Ile Pro
                275                 280                 285
Pro Ser Leu Asn Pro Ile Ile Tyr Gly Val Arg Thr Lys Gln Ile
                290                 295                 300
Arg Glu Arg Val Leu Tyr Val Phe Thr Lys Lys
                305                 310
```

<210> 11
<211> 354
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472434CD1

<400> 11

```
Met Arg Ser Leu Lys Ala Gly Gly Lys Gln Thr Val Tyr Val Ala
  1                 5                  10                  15
Gly Glu Gln Glu Ala Gly Ile Pro Asp Ala Gly Leu Ser Arg Gly
                 20                  25                  30
Glu Val Arg Ala Ala Leu His Gly Asp Gly Gly His Leu Gly Glu
                 35                  40                  45
Thr Thr Ala Ser Pro Thr Ala Pro Phe Ala Lys Leu Val Thr Thr
                 50                  55                  60
Asp Arg Thr Ser Thr Arg Phe Val Pro Gly Phe Pro Pro Arg Val
                 65                  70                  75
Thr Ser Leu Ser Val Ser Phe Leu Leu Gln Ser Asn Met Glu Ala
                 80                  85                  90
Arg Asn Asn Leu Ser Leu Met Asp Ile Cys Gly Thr Ser Ser Phe
                 95                 100                 105
Val Pro Leu Met Leu Asp Asn Phe Leu Glu Thr Gln Arg Thr Ile
                110                 115                 120
Ser Phe Pro Gly Cys Ala Leu Gln Met Tyr Leu Thr Leu Ala Leu
                125                 130                 135
Gly Ser Thr Glu Cys Leu Leu Leu Ala Val Met Ala Tyr Asp Arg
                140                 145                 150
Tyr Val Ala Ile Cys Gln Pro Leu Arg Tyr Pro Glu Leu Met Ser
                155                 160                 165
Gly Gln Thr Cys Met Gln Met Ala Ala Leu Ser Trp Gly Thr Gly
                170                 175                 180
Phe Ala Asn Ser Leu Leu Gln Ser Ile Leu Val Trp His Leu Pro
                185                 190                 195
Phe Cys Gly His Val Ile Asn Tyr Phe Tyr Glu Ile Leu Ala Val
                200                 205                 210
Leu Lys Leu Ala Cys Gly Asp Ile Ser Leu Asn Ala Leu Ala Leu
                215                 220                 225
Met Val Ala Thr Ala Val Leu Thr Leu Ala Pro Leu Leu Leu Ile
                230                 235                 240
Cys Leu Ser Tyr Leu Phe Ile Leu Ser Ala Ile Leu Arg Val Pro
                245                 250                 255
Ser Ala Ala Gly Arg Cys Lys Ala Phe Ser Thr Cys Ser Ala His
                260                 265                 270
Arg Thr Val Val Val Val Phe Tyr Gly Thr Ile Ser Phe Met Tyr
                275                 280                 285
Phe Lys Pro Lys Ala Lys Asp Pro Asn Val Asp Lys Thr Val Ala
```

```
                290                 295                 300
Leu Phe Tyr Gly Val Val Thr Pro Ser Leu Asn Pro Ile Ile Tyr
                305                 310                 315
Ser Leu Arg Asn Ala Glu Val Lys Ala Ala Val Leu Thr Leu Leu
                320                 325                 330
Arg Gly Gly Leu Leu Ser Arg Lys Ala Ser His Cys Tyr Cys Cys
                335                 340                 345
Pro Leu Pro Leu Ser Ala Gly Ile Gly
                350
```

<210> 12
<211> 319
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472435CD1

<400> 12

```
Met Glu Lys Ala Asn Glu Thr Ser Pro Val Met Gly Phe Val Leu
  1                 5                  10                  15
Leu Arg Leu Ser Ala His Pro Glu Leu Glu Lys Thr Phe Phe Val
                 20                  25                  30
Leu Ile Leu Leu Met Tyr Leu Val Ile Leu Leu Gly Asn Gly Val
                 35                  40                  45
Leu Ile Leu Val Thr Ile Leu Asp Ser Arg Leu His Thr Pro Met
                 50                  55                  60
Tyr Phe Phe Leu Gly Asn Leu Ser Phe Leu Asp Ile Cys Phe Thr
                 65                  70                  75
Thr Ser Ser Val Pro Leu Val Leu Asp Ser Phe Leu Thr Pro Gln
                 80                  85                  90
Glu Thr Ile Ser Phe Ser Ala Cys Ala Val Gln Met Ala Leu Ser
                 95                 100                 105
Phe Ala Met Ala Gly Thr Glu Cys Leu Leu Leu Ser Met Met Ala
                110                 115                 120
Phe Asp Arg Tyr Val Ala Ile Cys Asn Pro Leu Arg Tyr Ser Val
                125                 130                 135
Ile Met Ser Lys Ala Ala Tyr Met Pro Met Ala Ala Ser Ser Trp
                140                 145                 150
Ala Ile Gly Gly Ala Ala Ser Val Val His Thr Ser Leu Ala Ile
                155                 160                 165
Gln Leu Pro Phe Cys Gly Asp Asn Val Ile Asn His Phe Thr Cys
                170                 175                 180
Glu Ile Leu Ala Val Leu Lys Leu Ala Cys Ala Asp Ile Ser Ile
                185                 190                 195
Asn Val Ile Ser Met Glu Val Thr Asn Val Ile Phe Leu Gly Val
                200                 205                 210
Pro Val Leu Phe Ile Ser Phe Ser Tyr Val Phe Ile Ile Thr Thr
                215                 220                 225
Ile Leu Arg Ile Pro Ser Ala Glu Gly Arg Lys Lys Val Phe Ser
                230                 235                 240
Thr Cys Ser Ala His Leu Thr Val Val Ile Val Phe Tyr Gly Thr
                245                 250                 255
Leu Phe Phe Met Tyr Gly Lys Pro Lys Ser Lys Asp Ser Met Gly
                260                 265                 270
Ala Asp Lys Glu Asp Leu Ser Asp Lys Leu Ile Pro Leu Phe Tyr
                275                 280                 285
Gly Val Val Thr Pro Met Leu Asn Pro Ile Ile Tyr Ser Leu Arg
                290                 295                 300
Asn Lys Asp Val Lys Ala Ala Val Arg Arg Leu Leu Arg Pro Lys
                305                 310                 315
Gly Phe Thr Gln
```

<210> 13
<211> 309

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472438CD1

<220>
<221> unsure
<222> 125-126
<223> unknown or other

<400> 13
Met Ala Ala Gly Asn His Ser Thr Val Thr Glu Phe Ile Leu Lys
1 5 10 15
Gly Leu Thr Lys Arg Ala Asp Leu Gln Leu Pro Leu Phe Leu Leu
20 25 30
Phe Leu Gly Ile Tyr Leu Val Thr Ile Val Gly Asn Leu Gly Met
35 40 45
Ile Thr Leu Ile Cys Leu Asn Ser Gln Leu His Thr Pro Met Tyr
50 55 60
Tyr Phe Leu Ser Asn Leu Ser Leu Met Asp Leu Cys Tyr Ser Ser
65 70 75
Val Ile Thr Pro Lys Met Leu Val Asn Phe Val Ser Glu Lys Asn
80 85 90
Ile Ile Ser Tyr Ala Gly Cys Met Ser Gln Leu Tyr Phe Phe Leu
95 100 105
Val Phe Val Ile Ala Glu Cys Tyr Met Leu Thr Val Met Ala Tyr
110 115 120
Asp Arg Tyr Val Xaa Xaa Cys His Pro Leu Leu Tyr Asn Ile Ile
125 130 135
Met Ser His His Thr Cys Leu Leu Leu Val Ala Val Val Tyr Ala
140 145 150
Ile Gly Leu Ile Gly Ser Thr Ile Glu Thr Gly Leu Met Leu Lys
155 160 165
Leu Pro Tyr Cys Glu His Leu Ile Ser His Tyr Phe Cys Asp Ile
170 175 180
Leu Pro Leu Met Lys Leu Ser Cys Ser Ser Thr Tyr Asp Val Glu
185 190 195
Met Thr Val Phe Phe Ser Ala Gly Phe Asn Ile Ile Val Thr Ser
200 205 210
Leu Thr Val Leu Val Ser Tyr Thr Phe Ile Leu Ser Ser Ile Leu
215 220 225
Gly Ile Ser Thr Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr Cys
230 235 240
Ser Ser His Leu Ala Ala Val Gly Met Phe Tyr Gly Ser Thr Ala
245 250 255
Phe Met Tyr Leu Lys Pro Ser Thr Ile Ser Ser Leu Thr Gln Glu
260 265 270
Asn Val Ala Ser Val Phe Tyr Thr Thr Val Ile Pro Met Leu Asn
275 280 285
Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Val Lys Ala Ala Val
290 295 300
Gln Lys Thr Leu Arg Gly Lys Leu Phe
305

<210> 14
<211> 310
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472439CD1

<400> 14
Met Ala Ala Lys Asn Ser Ser Val Thr Glu Phe Ile Leu Glu Gly

```
  1               5                  10                  15
Leu Thr His Gln Pro Gly Leu Arg Ile Pro Leu Phe Phe Leu Phe
                20                  25                  30
Leu Gly Phe Tyr Thr Val Thr Val Val Gly Asn Leu Gly Leu Ile
                35                  40                  45
Thr Leu Ile Gly Leu Asn Ser His Leu His Thr Pro Met Tyr Phe
                50                  55                  60
Phe Leu Phe Asn Leu Ser Leu Ile Asp Phe Cys Phe Ser Thr Thr
                65                  70                  75
Ile Thr Pro Lys Met Leu Met Ser Phe Val Ser Arg Lys Asn Ile
                80                  85                  90
Ile Ser Phe Thr Gly Cys Met Thr Gln Leu Phe Phe Phe Cys Phe
                95                 100                 105
Phe Val Val Ser Glu Ser Phe Ile Leu Ser Ala Met Ala Tyr Asp
               110                 115                 120
Arg Tyr Val Ala Ile Cys Asn Pro Leu Leu Tyr Thr Val Thr Met
               125                 130                 135
Ser Cys Gln Val Cys Leu Leu Leu Leu Leu Gly Ala Tyr Gly Met
               140                 145                 150
Gly Phe Ala Gly Ala Met Ala His Thr Gly Ser Ile Met Asn Leu
               155                 160                 165
Thr Phe Cys Ala Asp Asn Leu Val Asn His Phe Met Cys Asp Ile
               170                 175                 180
Leu Pro Leu Leu Glu Leu Ser Cys Asn Ser Ser Tyr Met Asn Glu
               185                 190                 195
Leu Val Val Phe Ile Val Val Ala Val Asp Val Gly Met Pro Ile
               200                 205                 210
Val Thr Val Phe Ile Ser Tyr Ala Leu Ile Leu Ser Ser Ile Leu
               215                 220                 225
His Asn Ser Ser Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr Cys
               230                 235                 240
Ser Ser His Ile Ile Val Val Ser Leu Phe Phe Gly Ser Gly Ala
               245                 250                 255
Phe Met Tyr Leu Lys Pro Leu Ser Ile Leu Pro Leu Glu Gln Gly
               260                 265                 270
Lys Val Ser Ser Leu Phe Tyr Thr Ile Ile Val Pro Val Leu Asn
               275                 280                 285
Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Val Ala Leu
               290                 295                 300
Arg Arg Thr Leu Gly Arg Lys Ile Phe Ser
               305                 310
```

<210> 15
<211> 311
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472440CD1

<400> 15

```
Met Ala Ala Glu Asn Ser Ser Phe Val Thr Gln Phe Ile Leu Ala
  1               5                  10                  15
Gly Leu Thr Asp Gln Pro Gly Val Gln Ile Pro Leu Phe Phe Leu
                20                  25                  30
Phe Leu Gly Phe Tyr Val Val Thr Val Val Gly Asn Leu Gly Leu
                35                  40                  45
Ile Thr Leu Ile Arg Leu Asn Ser His Leu His Thr Pro Met Tyr
                50                  55                  60
Phe Phe Leu Tyr Asn Leu Ser Phe Ile Asp Phe Cys Tyr Ser Ser
                65                  70                  75
Val Ile Thr Pro Lys Met Leu Met Ser Phe Val Leu Lys Lys Asn
                80                  85                  90
Ser Ile Ser Tyr Ala Gly Cys Met Thr Gln Leu Phe Phe Phe Leu
                95                 100                 105
Phe Phe Val Val Ser Glu Ser Phe Ile Leu Ser Ala Met Ala Tyr
```

```
                  110                 115                 120
Asp Arg Tyr Val Ala Ile Cys Asn Pro Leu Leu Tyr Met Val Thr
                  125                 130                 135
Met Ser Pro Gln Val Cys Phe Leu Leu Leu Leu Gly Val Tyr Gly
                  140                 145                 150
Met Gly Phe Ala Gly Ala Met Ala His Thr Ala Cys Met Met Gly
                  155                 160                 165
Val Thr Phe Cys Ala Asn Asn Leu Val Asn His Tyr Met Cys Asp
                  170                 175                 180
Ile Leu Pro Leu Leu Glu Cys Ala Cys Thr Ser Thr Tyr Val Asn
                  185                 190                 195
Glu Leu Val Val Phe Val Val Val Gly Ile Asp Ile Gly Val Pro
                  200                 205                 210
Thr Val Thr Ile Phe Ile Ser Tyr Ala Leu Ile Leu Ser Ser Ile
                  215                 220                 225
Phe His Ile Asp Ser Thr Glu Gly Arg Ser Lys Ala Phe Ser Thr
                  230                 235                 240
Cys Ser Ser His Ile Ile Ala Val Ser Leu Phe Phe Gly Ser Gly
                  245                 250                 255
Ala Phe Met Tyr Leu Lys Pro Phe Ser Leu Leu Ala Met Asn Gln
                  260                 265                 270
Gly Lys Val Ser Ser Leu Phe Tyr Thr Thr Val Val Pro Met Leu
                  275                 280                 285
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Val Ala
                  290                 295                 300
Leu Lys Lys Ile Leu Asn Lys Asn Ala Phe Ser
                  305                 310
```

<210> 16
<211> 314
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472443CD1

<400> 16

```
Met Ala Lys Asn Asn Leu Thr Arg Val Thr Glu Phe Ile Leu Met
  1               5                  10                  15
Gly Phe Met Asp His Pro Lys Leu Glu Ile Pro Leu Phe Leu Val
                 20                  25                  30
Phe Leu Ser Phe Tyr Leu Val Thr Leu Leu Gly Asn Val Gly Met
                 35                  40                  45
Ile Met Leu Ile Gln Val Asp Val Lys Leu Tyr Thr Pro Met Tyr
                 50                  55                  60
Phe Phe Leu Ser His Leu Ser Leu Leu Asp Ala Cys Tyr Thr Ser
                 65                  70                  75
Val Ile Thr Pro Gln Ile Leu Ala Thr Leu Ala Thr Gly Lys Thr
                 80                  85                  90
Val Ile Ser Tyr Gly His Cys Ala Ala Gln Phe Phe Leu Phe Thr
                 95                 100                 105
Ile Cys Ala Gly Thr Glu Cys Phe Leu Leu Ala Val Met Ala Tyr
                110                 115                 120
Asp Arg Tyr Ala Ala Ile Arg Asn Pro Leu Leu Tyr Thr Val Ala
                125                 130                 135
Met Asn Pro Arg Leu Cys Trp Ser Leu Val Val Gly Ala Tyr Val
                140                 145                 150
Cys Gly Val Ser Gly Ala Ile Leu Arg Thr Thr Cys Thr Phe Thr
                155                 160                 165
Leu Ser Phe Cys Lys Asp Asn Gln Ile Asn Phe Phe Phe Cys Asp
                170                 175                 180
Leu Pro Pro Leu Leu Lys Leu Ala Cys Ser Asp Thr Ala Asn Ile
                185                 190                 195
Glu Ile Val Ile Ile Phe Phe Gly Asn Phe Val Ile Leu Ala Asn
                200                 205                 210
Ala Ser Val Ile Leu Ile Ser Tyr Leu Leu Ile Ile Lys Thr Ile
```

```
                215                 220                 225
Leu Lys Val Lys Ser Ser Gly Gly Arg Ala Lys Thr Phe Ser Thr
                230                 235                 240
Cys Ala Ser His Ile Thr Ala Val Ala Leu Phe Phe Gly Ala Leu
                245                 250                 255
Ile Phe Met Tyr Leu Gln Ser Gly Ser Gly Lys Ser Leu Glu Glu
                260                 265                 270
Asp Lys Val Val Ser Val Phe Tyr Thr Val Val Ile Pro Met Leu
                275                 280                 285
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Asp Ala
                290                 295                 300
Phe Arg Lys Val Ala Arg Arg Leu Gln Val Ser Leu Ser Met
                305                 310
```

<210> 17
<211> 346
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472445CD1

<400> 17

```
Met Asn Ala Ala Gln Ser His Tyr Pro Lys Arg Thr Asn Ala Gly
  1                 5                  10                  15
Thr Gly Asn Gln Ile Ser His Val Leu Thr Cys Lys Gln Ala Lys
                 20                  25                  30
Ile Ser Met Gly Glu Glu Asn Gln Thr Phe Val Ser Lys Phe Ile
                 35                  40                  45
Phe Leu Gly Leu Ser Gln Asp Leu Gln Thr Gln Ile Leu Leu Phe
                 50                  55                  60
Ile Leu Phe Leu Ile Ile Tyr Leu Leu Thr Val Leu Gly Asn Gln
                 65                  70                  75
Leu Ile Ile Ile Leu Ile Phe Leu Asp Ser Arg Leu His Thr Pro
                 80                  85                  90
Met Tyr Phe Phe Leu Arg Asn Leu Ser Phe Ala Asp Leu Cys Phe
                 95                 100                 105
Ser Thr Ser Ile Val Pro Gln Val Leu Val His Phe Leu Val Lys
                110                 115                 120
Arg Lys Thr Ile Ser Phe Tyr Gly Cys Met Thr Gln Ile Ile Val
                125                 130                 135
Phe Leu Leu Val Gly Cys Thr Glu Cys Ala Leu Leu Ala Val Met
                140                 145                 150
Ser Tyr Asp Arg Tyr Val Ala Val Cys Lys Pro Leu Tyr Tyr Ser
                155                 160                 165
Thr Ile Met Thr Gln Arg Val Cys Leu Trp Leu Ser Phe Arg Ser
                170                 175                 180
Trp Ala Ser Gly Ala Leu Val Ser Leu Val Asp Thr Ser Phe Thr
                185                 190                 195
Phe His Leu Pro Tyr Trp Gly Gln Asn Ile Ile Asn His Tyr Phe
                200                 205                 210
Cys Glu Pro Pro Ala Leu Leu Lys Leu Ala Ser Ile Asp Thr Tyr
                215                 220                 225
Ser Thr Glu Met Ala Ile Phe Ser Met Gly Val Val Ile Leu Leu
                230                 235                 240
Ala Pro Val Ser Leu Ile Leu Gly Ser Tyr Trp Asn Ile Ile Ser
                245                 250                 255
Thr Val Ile Gln Met Gln Ser Gly Glu Gly Arg Leu Lys Ala Phe
                260                 265                 270
Ser Thr Cys Gly Ser His Leu Ile Val Val Val Leu Phe Tyr Gly
                275                 280                 285
Ser Gly Ile Phe Thr Tyr Met Arg Pro Asn Ser Lys Thr Thr Lys
                290                 295                 300
Glu Leu Asp Lys Met Ile Ser Val Phe Tyr Thr Ala Val Thr Pro
                305                 310                 315
Met Leu Asn Pro Ile Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys
```

```
                320                 325                 330
Gly Ala Leu Arg Lys Leu Val Gly Arg Lys Cys Phe Ser His Arg
                335                 340                 345

Gln


<210> 18
<211> 316
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472446CD1

<400> 18
Met Glu Leu Trp Asn Phe Thr Leu Gly Ser Gly Phe Ile Leu Val
  1               5                  10                  15
Gly Ile Leu Asn Asp Ser Gly Ser Pro Glu Leu Leu Cys Ala Thr
                20                  25                  30
Ile Thr Ile Leu Tyr Leu Leu Ala Leu Ile Ser Asn Gly Leu Leu
                35                  40                  45
Leu Leu Ala Ile Thr Met Glu Ala Arg Leu His Met Pro Met Tyr
                50                  55                  60
Leu Leu Leu Gly Gln Leu Ser Leu Met Asp Leu Leu Phe Thr Ser
                65                  70                  75
Val Val Thr Pro Lys Ala Leu Ala Asp Phe Leu Arg Arg Glu Asn
                80                  85                  90
Thr Ile Ser Phe Gly Gly Cys Ala Leu Gln Met Phe Leu Ala Leu
                95                 100                 105
Thr Met Gly Gly Ala Glu Asp Leu Leu Leu Ala Phe Met Ala Tyr
               110                 115                 120
Asp Arg Tyr Val Ala Ile Cys His Pro Leu Thr Tyr Met Thr Leu
               125                 130                 135
Met Ser Ser Arg Ala Cys Trp Leu Met Val Ala Thr Ser Trp Ile
               140                 145                 150
Leu Ala Ser Leu Ser Ala Leu Ile Tyr Thr Val Tyr Thr Met His
               155                 160                 165
Tyr Pro Phe Cys Arg Ala Gln Glu Ile Arg His Leu Leu Cys Glu
               170                 175                 180
Ile Pro His Leu Leu Lys Val Ala Cys Ala Asp Thr Ser Arg Tyr
               185                 190                 195
Glu Leu Met Val Tyr Val Met Gly Val Thr Phe Leu Ile Pro Ser
               200                 205                 210
Leu Ala Ala Ile Leu Ala Ser Tyr Thr Gln Ile Leu Leu Thr Val
               215                 220                 225
Leu His Met Pro Ser Asn Glu Gly Arg Lys Lys Ala Leu Val Thr
               230                 235                 240
Cys Ser Ser His Leu Thr Val Val Gly Met Phe Tyr Gly Ala Ala
               245                 250                 255
Thr Phe Met Tyr Val Leu Pro Ser Ser Phe His Ser Thr Arg Gln
               260                 265                 270
Asp Asn Ile Ile Ser Val Phe Tyr Thr Ile Val Thr Pro Ala Leu
               275                 280                 285
Asn Pro Leu Ile Tyr Ser Leu Arg Asn Lys Glu Val Met Arg Ala
               290                 295                 300
Leu Arg Arg Val Leu Gly Lys Tyr Met Leu Pro Ala His Ser Thr
               305                 310                 315
Leu


<210> 19
<211> 453
<212> PRT
<213> Homo sapiens

<220>
```

<221> misc_feature
<223> Incyte ID No: 7472451CD1

<400> 19

```
Met Leu Tyr Lys Tyr Leu Glu Arg Asp Val Asn Ser Lys Glu Leu
  1               5                  10                  15
Gln Ser Gly Asn Gln Thr Ser Val Ser His Phe Ile Leu Val Gly
             20                  25                  30
Leu His His Pro Pro Gln Leu Gly Ala Pro Leu Phe Leu Ala Phe
         35                  40                  45
Leu Val Ile Tyr Leu Leu Thr Val Ser Gly Asn Gly Leu Ile Ile
     50                  55                  60
Leu Thr Val Leu Val Asp Ile Arg Leu His Arg Pro Met Cys Leu
 65                  70                  75
Phe Leu Cys His Leu Ser Phe Leu Asp Met Thr Ile Ser Cys Ala
             80                  85                  90
Ile Val Pro Lys Met Leu Ala Gly Phe Leu Leu Gly Ser Arg Ile
         95                 100                 105
Ile Ser Phe Gly Gly Cys Val Ile Gln Leu Phe Ser Phe His Phe
    110                 115                 120
Leu Gly Cys Thr Glu Cys Phe Leu Tyr Thr Leu Met Ala Tyr Asp
    125                 130                 135
Arg Phe Leu Ala Ile Cys Lys Pro Leu His Tyr Ala Thr Ile Met
    140                 145                 150
Thr His Arg Val Cys Asn Ser Leu Ala Leu Gly Thr Trp Leu Gly
    155                 160                 165
Gly Thr Ile His Ser Leu Phe Gln Thr Ser Phe Val Phe Arg Leu
    170                 175                 180
Pro Phe Cys Gly Pro Asn Arg Val Asp Tyr Ile Phe Cys Asp Ile
    185                 190                 195
Pro Ala Met Leu Arg Leu Ala Cys Ala Asp Thr Ala Ile Asn Glu
    200                 205                 210
Leu Val Thr Phe Ala Asp Ile Gly Phe Leu Ala Leu Thr Cys Phe
    215                 220                 225
Met Leu Ile Leu Thr Ser Tyr Gly Tyr Ile Val Ala Ala Ile Leu
    230                 235                 240
Arg Ile Pro Ser Ala Asp Gly Arg Arg Asn Ala Phe Ser Thr Cys
    245                 250                 255
Ala Ala His Leu Thr Val Val Ile Val Tyr Tyr Val Pro Cys Thr
    260                 265                 270
Phe Ile Tyr Leu Arg Pro Cys Ser Gln Glu Pro Leu Asp Gly Val
    275                 280                 285
Val Ala Val Phe Tyr Thr Val Ile Thr Pro Leu Leu Asn Ser Ile
    290                 295                 300
Ile Tyr Thr Leu Cys Asn Lys Glu Met Lys Ala Ala Leu Gln Arg
    305                 310                 315
Leu Gly Gly His Lys Glu Glu Val Glu Glu Ile Glu Leu Gly His
    320                 325                 330
Thr Thr Val Glu Gly His Ser Leu Ala Thr Gln Gly Gln Gln Gly
    335                 340                 345
Pro Arg His Phe Gly His His Asp Ser Glu Glu Pro Gln Val Asn
    350                 355                 360
Glu Gly Gln Ile Gly Glu Glu Val Val Leu Gly Gly Val Glu Val
    365                 370                 375
Arg Val His Pro Asp His Gln Gln Asp Glu Glu Val Pro Gln His
    380                 385                 390
Asn Leu Lys Lys Asn Tyr Met Thr Ile Ser Ala Asp Gly Glu Lys
    395                 400                 405
Ala Leu Asp Asn Ile Arg His Pro Leu Ile Lys Thr Leu Asn Asn
    410                 415                 420
Leu Glu Val Lys Gly Asp Phe Leu Asn Leu Met Lys Asp Val Tyr
    425                 430                 435
Glu Asn Pro Thr Pro Asn Leu Ser Lys Tyr Pro Glu Lys Leu Asn
    440                 445                 450
Ala Phe Pro
```

<210> 20
<211> 323
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472456CD1

<400> 20
Met Asn Pro Glu Asn Trp Thr Gln Val Thr Ser Phe Val Leu Leu
1 5 10 15
Gly Phe Pro Ser Ser His Leu Ile Gln Phe Leu Val Phe Leu Gly
20 25 30
Leu Met Val Thr Tyr Ile Val Thr Ala Thr Gly Lys Leu Leu Ile
35 40 45
Ile Val Leu Ser Trp Ile Asp Gln Arg Leu His Ile Gln Met Tyr
50 55 60
Phe Phe Leu Arg Asn Phe Ser Phe Leu Glu Leu Leu Leu Val Thr
65 70 75
Val Val Val Pro Lys Met Leu Val Val Ile Leu Thr Gly Asp His
80 85 90
Thr Ile Ser Phe Val Ser Cys Ile Ile Gln Ser Tyr Leu Tyr Phe
95 100 105
Phe Leu Gly Thr Thr Asp Phe Phe Leu Leu Ala Val Met Ser Leu
110 115 120
Asp Arg Tyr Leu Ala Ile Cys Arg Pro Leu Arg Tyr Glu Thr Leu
125 130 135
Met Asn Gly His Val Cys Ser Gln Leu Val Leu Ala Ser Trp Leu
140 145 150
Ala Gly Phe Leu Trp Val Leu Cys Pro Thr Val Leu Met Ala Ser
155 160 165
Leu Pro Phe Cys Gly Pro Asn Gly Ile Asp His Phe Phe Arg Asp
170 175 180
Ser Trp Pro Leu Leu Arg Leu Ser Cys Gly Asp Thr His Leu Leu
185 190 195
Lys Leu Val Ala Phe Met Leu Ser Thr Leu Val Leu Leu Gly Ser
200 205 210
Leu Ala Leu Thr Ser Val Ser Tyr Ala Cys Ile Leu Ala Thr Val
215 220 225
Leu Arg Ala Pro Thr Ala Ala Glu Arg Arg Lys Ala Phe Ser Thr
230 235 240
Cys Ala Ser His Leu Thr Val Val Val Ile Ile Tyr Gly Ser Ser
245 250 255
Ile Phe Leu Tyr Ile Arg Met Ser Glu Ala Gln Ser Lys Leu Leu
260 265 270
Asn Lys Gly Ala Ser Val Leu Ser Cys Ile Ile Thr Pro Leu Leu
275 280 285
Asn Pro Phe Ile Phe Thr Leu Arg Asn Asp Lys Val Gln Gln Ala
290 295 300
Leu Arg Glu Ala Leu Gly Trp Pro Arg Leu Thr Ala Val Met Lys
305 310 315
Leu Arg Val Thr Ser Gln Arg Lys
320

<210> 21
<211> 318
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472457CD1

<400> 21
Met Asn Pro Ala Asn His Ser Gln Val Ala Gly Phe Val Leu Leu
1 5 10 15

```
Gly Leu Ser Gln Val Trp Glu Leu Arg Phe Val Phe Phe Thr Val
                20                  25                  30
Phe Ser Ala Val Tyr Phe Met Thr Val Val Gly Asn Leu Leu Ile
                35                  40                  45
Val Val Ile Val Thr Ser Asp Pro His Leu His Thr Thr Met Tyr
                50                  55                  60
Phe Leu Leu Gly Asn Leu Ser Phe Leu Asp Phe Cys Tyr Ser Ser
                65                  70                  75
Ile Thr Ala Pro Arg Met Leu Val Asp Leu Leu Ser Gly Asn Pro
                80                  85                  90
Thr Ile Ser Phe Gly Gly Cys Leu Thr Gln Leu Phe Phe Phe His
                95                  100                 105
Phe Ile Gly Gly Ile Lys Ile Phe Leu Leu Thr Val Met Ala Tyr
                110                 115                 120
Asp Arg Tyr Ile Ala Ile Ser Gln Pro Leu His Tyr Thr Leu Ile
                125                 130                 135
Met Asn Gln Thr Val Cys Ala Leu Leu Met Ala Ala Ser Trp Val
                140                 145                 150
Gly Gly Phe Ile His Ser Ile Val Gln Ile Ala Leu Thr Ile Gln
                155                 160                 165
Leu Pro Phe Cys Gly Pro Asp Lys Leu Asp Asn Phe Tyr Cys Asp
                170                 175                 180
Val Pro Gln Leu Ile Lys Leu Ala Cys Thr Asp Thr Phe Val Leu
                185                 190                 195
Glu Leu Leu Met Val Ser Asn Asn Gly Leu Val Thr Leu Met Cys
                200                 205                 210
Phe Leu Val Leu Leu Gly Ser Tyr Thr Ala Leu Leu Val Met Leu
                215                 220                 225
Arg Ser His Ser Arg Glu Gly Arg Ser Lys Ala Leu Ser Thr Cys
                230                 235                 240
Ala Ser His Ile Ala Val Val Thr Leu Ile Phe Val Pro Cys Ile
                245                 250                 255
Tyr Val Tyr Thr Arg Pro Phe Arg Thr Phe Pro Met Asp Lys Ala
                260                 265                 270
Val Ser Val Leu Tyr Thr Ile Val Thr Pro Met Leu Asn Pro Ala
                275                 280                 285
Ile Tyr Thr Leu Arg Asn Lys Glu Val Ile Met Ala Met Lys Lys
                290                 295                 300
Leu Trp Arg Arg Lys Lys Asp Pro Ile Gly Pro Leu Glu His Arg
                305                 310                 315
Pro Leu His
```

<210> 22
<211> 1348
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 536482CB1

<400> 22

```
gaagccgcag tcccgcccgg cttcccggga cgcacagggc aggcgctggg catgcgcatg 60
ctcggcaggc ggggcagctg gagtctccag gcgctgcaat ctgttgcgcc gccgcccgcg 120
ggagccggca gatccgggtc ttgtggctaa gagtgacgtg gtcactcgaa tcaaaacaga 180
ggaggggggag gaagccggcg gccagaaacg gcagtggcag cagcgtccgg agcagccgca 240
gccttctgga agctccaggc ggtctttctg ccgagcctcg gtcccggccc ccatcctccc 300
cgccccatcg gttgttgtct gggcggattt aaacagtcaa gtaaaatcaa gctgggtaat 360
catggcagaa ggtggatttg atccctgtga atgtgtttgc tctcatgaac atgcaatgag 420
aagactgatc aatctgttac ggcagtccca gtcctactgc acagacacag agtgtcttca 480
ggaattaccg ggaccctctg gtgataatgg catcagtgtt acaatgatct tggtagcctg 540
gatggttatt gcattgatct tgttcttact gagacctcct aatctaagag gatccagcct 600
acctggaaag ccaaccagtc ctcataatgg acaagatcca ccagctcctc ctgtggacta 660
actttgtgat atgggaagtg aaaatagtta acaccttgca cgaccaaacg aacgaagatg 720
accagagtac tcttaacccc attagaactg tttttccttt tgtatctgca atatgggatg 780
gtattgtttt catgagcttc tagaaatttc acttgcaagt ttatttttgc ttcctgtgtt 840
```

actgccattc ctatttacag tatatttgag tgaatgatta tatttttaaa aagttacatg 900
gggctttttt ggttgtccta aacttacaaa cattccactc attctgtttg taactgtgat 960
tataattttt gtgataattt ctggcctgat tgaaggaaat ttgagaggtc tgcatttata 1020
tattttaaat agatttgata ggtttttaaa ttgctttttt tcataaggta tttataaagt 1080
tatttggggt tgtctgggat tgtgtgaaag aaaattagaa ccacgctgta tttacattta 1140
ccttggtagt ttatttgtgg atggcagttt tctgtagttt tggggactgt ggtagctctt 1200
ggattgtttt gcaaattaca gctgaaatct gtgtcatgga ttaaactggc ttatgtggct 1260
agaataggaa gagagaaaaa atgaaatggt tgtttactaa ttttatactc ccattaaaaa 1320
tctctaatgt taagaaaacc ttaaataa 1348

<210> 23
<211> 1446
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1316020CB1

<400> 23
gaaaggcacc agtcctaagg tgaacattaa gtgagatgat tctagttaca gacttagaac 60
aatttccagc acatagttaa atatccagga aattctggta ctgttatgtg tgggtgagct 120
gacctggatg tagatgtttt cctctctctt gctgacccct ccgccagttt tgtcttgtga 180
tgccattaac acatctctcc ctttctgacc tggctcctgc ccattggtgt cccaagaaat 240
cgtgagaata gttagccccc cgtctcccca gcctgttgct ttctcgtgta gttgttcaca 300
gtagttgaga agttgaagag cttttgccta ttgaaggtgc actgagaata aactctttcc 360
tgccaccaga attgcagtgg ttcacggcct gcactcattc ccatgaatgc agttaatagc 420
cacagaaatg tcacattaag caaagcagcc agggtctcat cgtgttgaga ctcgagtctc 480
tcagaccttg gattcattcc ctggtgtctt tgagcctcag tttcctcatt ggtaaaagag 540
aagtgaagca gtgtctcaca gggtcattac agagattaaa tgaaataaat gaaataacat 600
agaccaggag ggcgtggtgt ttaaaagtca cagatggggc accctcgggc catccagccc 660
agtgtttttct ttagccccta tgatgttcat tttttgttat atcccattag gtgcccatat 720
ttaaaaattg ggagatttca cataaaatta aaaggtctgc attttctttt ttcttttctt 780
ttttttttttt ttgagacaca gtctcactct gtcaccaggc tagagtgcag tggcacgatc 840
tcagctcact gcaacctctg cctcccaggt tcaagtaatt ctcctgcctc agcctcccaa 900
gtagctggga ctacaggcac gtgccaccac gcccagctaa tttttgtatt tttagcagag 960
atggggtttc accacattgg ccaggatggt ctcgatctca acctcgtgat ccacccacct 1020
cggtctccca aagcgctggg attacaggcg tgagccaccg cgccaagcca aggtctgcat 1080
ttttctttag aactcagaac acccaatagt cctaggcccc catcctcgca tggcagcaag 1140
ctaaataagc atcttcccac tgcgagttgg ggcatgaccc agcctatggt ttgccatact 1200
ccctcttttt ctccgttttt tcattaattg tgaacctgac ctgcatcacc ctttcatgtc 1260
agtgctctcc aaacctgctt gcttgcaccc ctctagtcga aatattttgt gcttacccca 1320
atatatgtgt gtgactattg aactctattc gtagactgct tgtactaatg tcatttgcat 1380
cataaaatat tcatatccaa taaacatatt aaaaggatga gataagaaac cgaaaaaaaa 1440
aaaaaa 1446

<210> 24
<211> 1463
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2816437CB1

<400> 24
cgagtggttc tcctgcctca gcctcccgaa tggctgggac tacaggcatg tgccaccatg 60
cttggctaat ttttgtattt ttagtaggga tgggctgtca ccatgttggc aaggatggtc 120
tcgatctctt gacctcatga tccgcccacc tgggattact tatatgaaaa taaaatttta 180
aataaaaaat agcatttgat tacaactatt ggtgagacta ttagtgtgaa gtcatatttt 240
tacttacatt gacaaaataa ccattctgta tattggatat tgacttctat tgacaaaata 300
gccataacaa tattctgatt tagaataata ctccttttct gctgtatatt tgcaagcttt 360
tatcaaatat tacgggagct caatagaaat caacaatatg aatctttatt taccacaaac 420
attattgatg cctatgcttg tttttcttaa aatctattag ctttttatga cacatttata 480
ctttttcagt tgtttattac tgagaagtgg ctgtcctgcc agaaaactgc tattctcagc 540
tgtatccaca atgactaata gtgagtggaa gtgcagtgga taaaagcaaa tgtgtcttct 600
ctgtattttt tttcatccat tggctaaaaa acaggaggat gcagaagatg aaagaaaata 660

caatccctga aagatcatga agaagctctc aatcagacaa aaaaccccat aggagatttt 720
ttttttttt cattattgta ctctctgctt ctatgagttt aaatttcttt tacacttcac 780
atagaaataa acctaggctt acatagaaag aaacctaggc ttttggccag gcatggtggc 840
ttaagcctgt aattccagca ctttggaggg ccgaggcggg cggatcatct gaggtcagga 900
gttcacgcaa atcaggggge ccgtgaaagg caactgagca gggatccatg ggaaagacac 960
cctcagaggc acaagattct ctcgttacct ttcagtttgc tgatacttca gttaaagtct 1020
cctgggaaac gtctgcatta ggttcttcct ctgtagttct tcttaccttg cctgtaaaac 1080
aaaacctatc tagtgtctgc ataggtttcc acttcttgtc cccacctgag gaatggaaag 1140
caacggcaca gtccttgctc atgttttgga gtgaaaggag cttgaaggtc atgtgagctt 1200
tgccaaggct tctcctggcc tcatgtcaga tacagctcct aactcccaag cagcctacca 1260
tagtgtcctc ctttttttgc gtgtgtgatg gggtttcgca cttgttgccc aggctggagt 1320
gcaatgggta cagtctcggc tcactgcaac ctccgcctcc caggttcaag tgattcttct 1380
gcctcagcct ctcaagtaac tgggattaca ggcatgcgcc actaagggac ggagaccact 1440
cctcatattg tcttatgccc aat 1463

<210> 25
<211> 1435
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2289894CB1

<400> 25
caacatcaga gtgacagcca gtaccatctc cgacaggggc tgagttgctg gagctgggct 60
ggggcagggg agaaagacag cactcatcct tgcaccctc catgggcctg gccaagcccc 120
caagaggatg gcagcctggg cgtcggagcc acctcctggg cagccaatga ggtgaggggc 180
cggaggagca agggacaaga ggagcagagg acaggtgatg gaaatcctgc agctttaggc 240
tccattctgc catctacatc ccagcgcagg gtgaagcctg agagcccaaa tggccaactc 300
cacagggctg aacgcctcag aagtcgcagg ctcgttgggg ttgatcctgg cagctgtcgt 360
ggaggtgggg gcactgctgg gcaacggcgc gctgctggtc gtggtgctgc gcacgccggg 420
actgcgcgac gcgctctacc tggcgcacct gtgcgtcgtg gacctgctgg cggccgcctc 480
catcatgccg ctgggcctgc tggccgcacc gccgcccggg ctgggccgcg tgcgcctggg 540
ccccgcgcca tgccgcgccg ctcgcttcct ctccgccgct ctgctgccgg cctgcacgct 600
cggggtggcc gcacttggcc tggcacgcta ccgcctcatc gtgcacccgc tgcggccagg 660
ctcgcggccg ccgcctgtgc tcgtgctcac cgccgtgtgg gccgcggcgg gactgctggg 720
cgcgctctcc ctgctcggcc cgccgcccgc accgcccctt gctcctgctc gctgctcggt 780
cctggctggg ggcctcgggc ccttccggcc gctctgggcc ctgctggcct tcgcgctgcc 840
cgccctcctg ctgctcggcg cctacggcgg catcttcgtg gtggcgcgtc gcgctgccct 900
gaggccccca cggccggcgc gcgggtcccg actccgctcg gactctctgg atagccgcct 960
ttccatcttg ccgccgctcc ggcctcgcct gcccgggggc aaggcggccc tggccccagc 1020
gctggccgtg ggccaatttg cagcctgctg gctgccttat ggctgcgcgt gcctggcgcc 1080
cgcagcgcgg gccgcggaag ccgaagcggc tgtcacctgg gtcgcctact cggccttcgc 1140
ggctcacccc ttcctgtacg ggctgctgca gcgccccgtg cgcttggcac tgggccgcct 1200
ctctcgccgt gcactgcctg gacctgtgcg ggcctgcact ccgcaagcct ggcacccgcg 1260
ggcactcttg caatgcctcc agagacccca agagggccct gccgtaggcc cttctgaggc 1320
tccagaacag acccccgagt tggcaggagg gcggagcccc gcataccagg ggccacctga 1380
gagttctctc tcctgagcag gagaaaggag ggtggtttcc gtgggggctc atcca 1435

<210> 26
<211> 2147
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7066050CB1

<400> 26
ctcggttcaa ggcagcgcga ctgcgggtgg cgcacgacca gggcgcagac cttggggcgc 60
gcggcccatg gagtcggggc tgctgcggcc ggcgccggtg agcgaggtca tcgtcctgca 120
ttacaactac accggcaagc tccgcggtgc gcgctaccag ccgggtgccg gcctgcgcgc 180
cgacgccgtg gtgtgcctgg cggtgtgcgc cttcatcgtg ctagagaatc tagccgtgtt 240
gttggtgctc ggacgccacc cgcgcttcca cgctcccatg ttcctgctcc tgggcagcct 300
cacgttgtcg gatctgctgg caggcgccgc ctacgccgcc aacatcctac tgtcggggcc 360
gctcacgctg aaactgtccc ccgcgctctg gttcgcacgg gagggaggcg tcttcgtggc 420

actcactgcg tccgtgctga gcctcctggc catcgcgctg gagcgcagcc tcaccatggc 480
gcgcaggggg cccgcgcccg tctccagtcg ggggcgcacg ctggcgatgg cagccgcggc 540
ctggggcgtg tcgctgctcc tcgggctcct gccagcgctg ggctggaatt gcctgggtcg 600
cctggacgct tgctccactg tcttgccgct ctacgccaag gcctacgtgc tcttctgcgt 660
gctcgccttc gtgggcatcc tggccgctat ctgtgcactc tacgcgcgca tctactgcca 720
ggtacgcgcc aacgcgcggc gcctgccggc acggcccggg actgcgggga ccacctcgac 780
ccgggcgcgt cgcaagccgc gctcgctggc cttgctgcgc acgctcagcg tggtgctcct 840
ggcctttgtg gcatgttggg gccccctctt cctgctgctg ttgctcgacg tggcgtgccc 900
ggcgcgcacc tgtcctgtac tcctgcaggc cgatcccttc ctgggactgg ccatggccaa 960
ctcacttctg aaccccatca tctacacgct caccaaccgc gacctgcgcc acgcgctcct 1020
gcgcctggtc tgctgcggac gccactcctg cggcagagac ccgagtggct cccagcagtc 1080
ggcgagcgcg gctgaggctt ccgggggcct gcgccgctgc ctgcccccgg gccttgatgg 1140
gagcttcagc ggctcggagc gctcatcgcc ccagcgcgac gggctggaca ccagcggctc 1200
cacaggcagc cccggtgcac ccacagccgc ccggactctg gtatcagaac cggctgcaga 1260
ctgacaccct cggcccacga ctgtcttccc aagttttaca gacttgttct ttttacataa 1320
aggaatttgt aggaaatgca gccaaaggtg cagtcggaaa agatgcaggg gaaatgtatt 1380
tatgcagcga caccccacaa tgtgaacaaa cagacaaaaa atctgtgccc tcgtggaatt 1440
gacgttctgc ttgggaacac agaaaagaac tcggtgatga aataatggag atgattccag 1500
tgacaaacga cagagatggt gatggtggtc agggaagacc tctctgcaga ggtagtgact 1560
tgtgatgtga gctgagacct ctgtcctggg aagaccaaaa gaaaagcatt tcaggatgag 1620
ggaatggcat gcgcaaaggc cctgaggctg aaatgtgccc atgtgttcta agaaatgcag 1680
cgatgctggt gtgcctggag cagggacgga gggggagaat gggaggagac aaggagctga 1740
aggagtagtt cccgaaggac cttgtgggtg atatagagga cttcgctttt gctctgagtg 1800
aggtgggagc catagaagct tctaagcaga agagggactt gccctaattc aggtgatcac 1860
aggtgtcttg tggcctccat gggaggttga aaaccagaga aggtgaaggg gggctgcact 1920
gagccacagg aacaatgatg gagattccag ctaagcccag accccgtgga ttctagatag 1980
attttagagg cagcagacag aattactgag gaattgagtg taagagtgga ataaagttat 2040
caaggacaat gccaagggtg gggcacccca aaatttgact ctgggagact cagccaaatc 2100
ctatctggta ataaaatttc ttttttattt ttaaaaaaaa aaaaaaa 2147

<210> 27
<211> 1989
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5376785CB1

<400> 27
gcctcagcaa tggcggacgc tcctccccca gcctcgctgc cgctgacctg attttccagt 60
gaggacattg tgcgctccct ctgcctgcca gacttccctg agccagctga cctcttctgg 120
aagtacctgg acttggccac cttcatcctg ctctacatcc tgcccctcct catcatctct 180
gtggcctacg ctcgtgtggc caagaaactg tggctgtgta atatgattgg cgatgtgacc 240
acagagcagt actttgccct gcggcgcaaa aagaagaaga ccatcaagat gttgatgctg 300
gtggtagtcc tctttgccct ctgctggttc cccctcaact gctacgtcct cctcctgtcc 360
agcaaggtca tccgcaccaa caatgccctc tactttgcct tccactggtt tgccatgagc 420
agcacctgct ataacccctt catatactgc tggctgaacg agaacttcag gattgagcta 480
aaggcattac tgagcatgtg tcaaagacct cccaagcctc aggaggacag gcaaccctcc 540
ccagttcctt ccttcagggt ggcctggaca gagaagaatg atggccagag ggctcccctt 600
gccaataacc tcctgcccac ctcccaactc cagtctggga agacagacct gtcatctgtg 660
gaacccattg tgacgatgag ttagaagagg ttgggaagag ggagtgggag gggtctgtct 720
ccacctgagg cagggaaaga gagcctattc tcacacatga tcttcagagt gctggaaaca 780
cactcctgca gaagctgtag gactcttgaa ttcctaggaa actgtccagc ctcctagccc 840
catgtgatgt gaaaactaaa aggcaccacc aactagacat gtgttcataa attcccatct 900
aagaaacact gggaggcaca gcagcctgta tctctgagga agaggagcga ggacaacgtt 960
ggcccagatg ggggctgaat cattcaactg cctccatctg tggggcagct gctgccttac 1020
agcccttcct actgctgagc atcccgaagg gagacctaaa tcatactttg ggtgtggtga 1080
cccagatgca cagagctctg cttgaaacag gtacacgggc cagggaaatg ccagcaagcc 1140
agagcgggcg tggagatttt tatgcctcac tttctggagt cactgggcca tgatgaatca 1200
taagtcttca gtggcctagc aatatccaga taagaaagga ccaacttggg ttccttaaaa 1260
caaagggaaa ttattattgc cacttagaaa aattcagaaa agcacacact cacatacaca 1320
cacacaaaat cactctctta tcccatccat ttgtgataac atctgtgaac atgctgtggc 1380
tctatttgca acattttcct tcgtgtgtgt gattgtgtgc atgtgtgcat gacctttttt 1440
ttttcttttt ttttttttg gagaagtgga aattcttcgc ttcgtgttcc cccgaggttg 1500
tggggttctc gggtggccac aactcttcgg cctcactgtg aaagctcgtg ccctcccaag 1560
gtttcgattg cggttccttc ctgtgccttc agccctcccc ttaggttacg ctggggcact 1620

taccgggtcg cccggccaca tattgcccgg gataagctct cttggcattt ttatagtgta 1680
agaagacccg gggtttctcc cccgggtgtt agcgccaggg attgggtctc agattttctt 1740
gtgaccttcg gttgattccc accttgcctt aggcgccttc cccaagagtg tgtgcgggga 1800
tttcacgggg cgttgagcca cccgccggcc gcggccactt tttcttccac ctctgtatct 1860
ctttcacatt cacgcttggc tcggtgatta cccagaccgg ggggtacctt tcccccgggt 1920
ggtcccccc aaggcttttg acgaccgggt gacagtcccc agggtgacct gtacccgaga 1980
aggattgcc 1989

<210> 28
<211> 942
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3082743CB1

<400> 28
atggacagtc taaaccaaac aagagtgact gaatttgtct tcttgggact cactgataac 60
cgggtgctgg aaatgctgtt tttcatggca ttctcagcca tttatatgct aacgctttca 120
gggaacattc tcatcatcat tgccacagtc tttactccaa gtctccatac ccccatgtat 180
ttcttcctga gcaatctgtc ctttattgac atctgccact catctgtcac tgtgcctaag 240
atgttggagg gttttgcttt agaaagaaag accatttcct ttgacaactg catcacacag 300
ctcttcttcc tacatctctt tgcctgtgcc gagatctttc tgctgatcat tgtggcgtat 360
gatcgttacg tggctatctg cactccactc cactacccca atgtgatgaa catgagagtc 420
tgtatacagc ttgtctttgc tctctggttg gggggtactg ttcactcact agggcagacc 480
ttcttgacta ttcgtctacc ttactgtggc cccaacatta ttgacagcta cttctgtgat 540
gtgcctcttg ttatcaagct ggcctgcaca gatacatacc tcacaggaat actgattgtg 600
accaatagtg gaaccatctc cctctcctgt ttcttggccg tggtcacctc ctatatggtc 660
atcctggttt ctcttcgaaa acactcagct gaagggcgcc agaaagccct gtctacctgc 720
tcggcccact tcatggtggt tgccctcttc tttgggccat gtatcttcat ctatactcgg 780
ccagacacca gcttctccat tgacaaggtg gtgtctgtct tctacacagt ggtcacccct 840
ttgctgaatc ccttcattta caccttgagg aatgaggagg taaaaagtgc catgaagcag 900
ctcaggcaga gacaagtttt tttcacgaaa tcatatacat aa 942

<210> 29
<211> 948
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472361CB1

<400> 29
atgggagact ggaataacag tgatgctgtg gagcccatat ttatcctgag gggttttcct 60
ggactggagt atgttcattc ttggctctcc atcctcttct gtcttgcata tttggtagca 120
tttatgggta atgttaccat cctgtctgtc atttggatag aatcctctct ccatcagccc 180
atgtattact ttatttccat cttagcagtg aatgacctgg ggatgtccct gtctacactt 240
cccaccatgc ttgctgtgtt atggttggat gctccagaga tccaggcaag tgcttgctat 300
gctcagctgt tcttcatcca cacattcaca ttcctggagt cctcagtgtt gctggccatg 360
gcctttgacc gttttgttgc tatctgccat ccactgcact accccaccat cctcaccaac 420
agtgtaattg gcaaaattgg tttggcctgt ttgctacgaa gcttgggagt tgtacttccc 480
acacctttgc tactgagaca ctatcactac tgccatggca atgccctctc tcacgccttc 540
tgtttgcacc aggatgttct aagattatcc tgtacagatg ccaggaccaa cagtatttat 600
gggctttgtg tagtcattgc cacactaggt gtggattcaa tcttcatact tctttcttat 660
gttctgattc ttaatactgt gctggatatt gcatctcgtg aagagcagct aaaggcactc 720
aacacatgtg tatcccatat ctgtgtggtg cttatcttct ttgtgccagt tattggggtg 780
tcaatggtcc atcgctttgg gaagcatctg tctcccatag tccacatcct catggcagac 840
atctaccttc ttcttccccc agtccttaac cctattgtct atagtgtcag aacaaagcag 900
attcgtctag gaattctcca caagtttgtc ctaaggagga ggttttaa 948

<210> 30
<211> 1071
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472363CB1

<400> 30
atgattcatg gaggagatcc aaacatcaac attaacagga gtttggaaga agcccattcc 60
aacctcatgg ataacgttga ggggttcaag acttcagtgg aggaagcagc tgcagatatg 120
gtggaaatag caagagaaat ggaattagaa gtgaaacctg aagatggaac tgaatgctgc 180
aatctcacga caaaaggtct ggaagacttc cacatgtgga tctccgggcc tttctgctct 240
gtttaccttg tggctttgct gggcaatgcc accattctgc tagtcatcaa ggtagaacag 300
actctccggg agcccatgtt ctacttcctg gccattcttt ccactattga tttggccctt 360
tctgcaacct ctgtgcctcg catgctgggt atcttctggt ttgatgctca cgagattaac 420
tatggagctt gtgtggccca gatgtttctg atccatgcct tcactggcat ggaggctgag 480
gtcttactgg ctatggcttt tgaccgttat gtggccatct gtgctccact acattacgca 540
accatcttga catccctagt gttggtgggc attagcatgt gcattgtaat tcgtcccgtt 600
ttacttacac ttcccatggt ctatcttatc taccgcctac ccttttgtca ggctcacata 660
atagcccatt cctactgtga gcacatgggc attgcaaaat tgtcctgtgg aaacattcgt 720
atcaatggta tctatgggct ttttgtagtt tctttctttg ttctgaacct ggtgctcatt 780
ggcatctcgt atgtttacat tctccgtgct gtcttccgcc tcccatcaca tgatgctcag 840
ctaaaagccc taagcacgtg tggcgctcat gttggagtca tctgtgtttt ctatatccct 900
tcagtcttct ctttccttac tcatcgattt ggacaccaaa taccaggtta cattcacatt 960
cttgttgcca atctctattt gattatccca ccctctctca accccatcat ttatggggtg 1020
aggaccaaac agattcgaga gcgagtgctc tatgttttta ctaaaaaata a 1071

<210> 31
<211> 1001
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472364CB1

<400> 31
actgtgattt ggaaaaatgt tttatcacaa caagagcata tttcacccag tcacattttt 60
cctcattgga atcccaggtc tggaagactt ccacatgtgg atctccgggc ctttctgctc 120
tgtttacctt gtggctttgc tgggcaatgc caccattctg ctagtcatca aggtagaaca 180
gactctccgg gagcccatgt tctacttcct ggccattctt tccactattg atttggccct 240
ttctgcaacc tctgtgcctc gcatgctggg tatcttctgg tttgatgctc acgagattaa 300
ctatggagct tgtgtggccc agatgtttct gatccatgcc ttcactggca tggaggctga 360
ggtcttactg gctatggctt ttgaccgtta tgtggccatc tgtgctccac tacattacgc 420
aaccatcttg acatccctag tgttggtggg cattagcatg tgcattgtaa ttcgtcccgt 480
tttacttaca cttcccatgg tctatcttat ctaccgccta ccctttttgtc aggctcacat 540
aatagcccat tcctactgtg agcacatggg cattgcaaaa ttgtcctgtg gaaacattcg 600
tatcaatggt atctatgggc tttttgtagt ttctttcttt gttctgaacc tggtgctcat 660
tggcatctcg tatgtttaca ttctccgtgc tgtcttccgc ctcccatcac atgatgctca 720
gctaaaagcc ctaagcacgt gtggcgctca tgttggagtc atctgtgttt tctatatccc 780
ttcagtcttc tctttcctta ctcatcgatt tggacaccaa ataccaggtt acattcacat 840
tcttgttgcc aatctctatt tgattatccc accctctctc aaccccatca tttatggggt 900
gaggaccaaa cagattcgag agcgagtgct ctatgttttt actaaaaaat aagactctta 960
ccatgttatt ttactaaggg ctttgatcct tctataaaga c 1001

<210> 32
<211> 1065
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472434CB1

<400> 32
atgaggtccc tgaaagcagg gggcaagcag actgtctatg tggcagggga gcaagaggca 60
ggaatacctg acgctggcct ttcccgagga gaagtgagag cagctctgca tggcgatgga 120
ggtcacctgg gagagaccac agcctcgccc accgctccct ttgcaaagct ggtcacaact 180
gaccgcacct ccaccagatt cgtgcctggc ttccctcctc gtgtgacatc attgtcagta 240
tcatttctcc tacaaagcaa tatggaggcc agaaacaacc tctccctcat ggacatctgc 300

ggcacctcct cctttgtgcc tctcatgcta gacaatttcc tggaaaccca gaggaccatt 360
tccttccctg gctgtgccct gcagatgtac ctgaccctgg cgctgggatc aacggagtgc 420
ctgctgctgg ctgtgatggc atatgaccgt tatgtggcta tctgccagcc gcttaggtac 480
ccagagctca tgagtgggca gacctgcatg cagatggcag cgctgagctg ggggacaggc 540
tttgccaact cactgctaca gtccatcctt gtctggcacc tccccttctg tggccacgtc 600
atcaactact tctatgagat cttggcagtg ctaaaactgg cctgtgggga catctccctc 660
aatgcgctgg cattaatggt ggccacagcc gtcctgacac tggcccccct cttgctcatc 720
tgcctgtctt acctttttcat cctgtctgcc atccttaggg taccctctgc tgcaggccgg 780
tgcaaagcct tctccacctg ctcagcccac cgcacagtgg tggtggtttt ttatgggaca 840
atctccttca tgtacttcaa acccaaggcc aaggatccca acgtggataa gactgtcgca 900
ttgttctacg gggttgtgac gccctcgctg aaccccatca tttacagcct gaggaatgca 960
gaggtgaaag ctgccgtcct aactctgctg agaggaggtt tgctctccag gaaagcatcc 1020
cactgctact gctgccctct gcccctgtca gctggcatag gctag 1065

<210> 33
<211> 963
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472435CB1

<400> 33
atggaaaaag ccaatgagac ctcccctgtg atggggttcg ttctcctgag gctctctgcc 60
cacccagagc tggaaaagac attcttcgtg ctcatcctgc tgatgtacct cgtgatcctg 120
ctgggcaatg gggtcctcat cctggtgacc atccttgact cccgcctgca cacgcccatg 180
tacttcttcc tagggaacct ctccttcctg gacatctgct tcactacctc ctcagtccca 240
ctggtcctgg acagcttttt gactccccag gaaaccatct ccttctcagc ctgtgctgtg 300
cagatggcac tctcctttgc catggcagga acagagtgct tgctcctgag catgatggca 360
tttgatcgct atgtggccat ctgcaacccc cttaggtact ccgtgatcat gagcaaggct 420
gcctacatgc ccatggctgc cagctcctgg gctattggtg gtgctgcttc cgtggtacac 480
acatccttgg caattcagct gcccttctgt ggagacaatg tcatcaacca cttcacctgt 540
gagattctgg ctgttctaaa gttggcctgt gctgacattt ccatcaatgt gatcagcatg 600
gaggtgacga atgtgatctt cctaggagtc ccggttctgt tcatctcttt ctcctatgtc 660
ttcatcatca ccaccatcct gaggatcccc tcagctgagg ggaggaaaaa ggtcttctcc 720
acctgctctg cccacctcac cgtggtgatc gtcttctacg ggaccttatt cttcatgtat 780
gggaagccta agtctaagga ctccatggga gcagacaaag aggatctttc agacaaactc 840
atccccctttt tctatggggt ggtgaccccg atgctcaacc ccatcatcta tagcctgagg 900
aacaaggatg tgaaggctgc tgtgaggaga ctgctgagac caaaaggctt cactcagtga 960
tgg 963

<210> 34
<211> 1101
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472438CB1

<220>
<221> unsure
<222> 545, 547
<223> a, t, c, g, or other

<400> 34
atgaactctc aaaacgaacc aaccaaaact ccctacaaga ataaactgga aggaataaaa 60
cccaaactaa gcagaataga aacattagag acaaaagaaa gaaacaatcc agataaaagc 120
agaaaagggg cagaaccaag agatctcagg tggcctccca cccagaggag aatggctgca 180
ggaaatcact ctacagtgac agagttcatt ctcaagggtt taacgaagag agcagacctc 240
cagctccccc tctttctcct cttcctcggg atctacttgg tcaccatcgt ggggaacctg 300
ggcatgatca ctctaatttg tctgaactct cagctgcaca cccccatgta ctactttctc 360
agcaatctgt cactcatgga tctctgctac tcctccgtca ttacccctaa gatgctggtg 420
aactttgtgt cagagaaaaa catcatctcc tacgcagggt gcatgtcaca gctctacttc 480
ttccttgttt ttgtcattgc tgagtgttac atgctgacag tgatggccta cgaccgctat 540
gttgncntct gccacccttt gctttacaac atcattatgt ctcatcacac ctgcctgctg 600

ctggtggctg tggtctacgc catcggactc attggctcca caatagaaac tggcctcatg 660
ttaaaactgc cctattgtga gcacctcatc agtcactact tctgtgacat cctccctctc 720
atgaagctgt cctgctctag cacctatgat gttgagatga cagtcttctt ttcggctgga 780
ttcaacatca tagtcacgag cttaacagtt cttgtttctt acaccttcat tctctccagc 840
atcctcggca tcagcaccac agaggggaga tccaaagcct tcagcacctg cagctcccac 900
cttgcagccg tgggaatgtt ctatggatca actgcattca tgtacttaaa accctccaca 960
atcagttcct tgacccagga gaatgtggcc tctgtgttct acaccacggt aatccccatg 1020
ttgaatcccc taatctacag cctgaggaac aaggaagtaa aggctgccgt gcagaaaacg 1080
ctgaggggta aactgttttg a 1101

<210> 35
<211> 933
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472439CB1

<400> 35
atggcagcca aaaactcttc tgtgacagag tttatcctcg aaggcttaac ccaccagccg 60
ggactgcgga tccccctctt cttcctgttt ctgggtttct acacggtcac cgtggtgggg 120
aacctgggct tgataaccct gattgggctg aactctcacc tgcacactcc catgtacttc 180
ttcctttta acctctcttt aatagatttc tgtttctcca ctaccatcac tcccaaaatg 240
ctgatgagtt ttgtctcaag gaagaacatc atttccttca cagggtgtat gactcagctc 300
ttcttcttct gcttctttgt cgtctctgag tccttcatcc tgtcagcgat ggcgtatgac 360
cgctacgtgg ccatctgtaa cccactgttg tacacagtca ccatgtcttg ccaggtgtgt 420
ttgctccttt tgttgggtgc ctatgggatg gggtttgctg gggccatggc ccacacagga 480
agcataatga acctgacctt ctgtgctgac aaccttgtca atcatttcat gtgtgacatc 540
cttcctctcc ttgagctctc ctgcaacagc tcttacatga atgagctggt ggtctttatt 600
gtggtggctg ttgacgttgg aatgcccatt gtcactgtct ttatttctta tgccctcatc 660
ctctccagca ttctacacaa cagttctaca gaaggcaggt ccaaagcctt tagtacttgc 720
agttcccaca taattgtagt ttctcttttc tttggttctg gtgctttcat gtatctcaaa 780
ccccttccca tcctgccccc cgagcaaggg aaagtgtcct ccctgttcta taccataata 840
gtccccgtgt taaacccatt aatctatagc ttgaggaaca aggatgtcaa agttgccctg 900
aggagaactt tgggcagaaa aatcttttct taa 933

<210> 36
<211> 936
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472440CB1

<400> 36
atggctgctg agaattcctc cttcgtgaca cagtttatcc tcgcaggctt aactgaccaa 60
ccgggagtcc agatcccccc cttcttcctg tttctaggct tctacgtggt cactgtggtg 120
gggaacctgg gcttgataac cctgataagg ctcaactctc acttgcacac ccctatgtac 180
ttcttcctct ataacttgtc cttcatagat ttctgctatt ccagtgttat cactcccaaa 240
atgctgatga gctttgtctt aaagaagaac agcatctcct acgcagggtg tatgactcag 300
ctcttcttct ttcttttctt tgttgtctct gagtccttca tcctgtcagc aatggcgtat 360
gaccgctatg tggccatctg taacccactg ttgtacatgg tcaccatgtc tccccaggtg 420
tgttttctcc ttttgttggg tgtctatggg atggggtttg ctggggccat ggcccacaca 480
gcgtgcatga tgggtgtgac cttctgtgcc aataaccttg tcaaccacta catgtgtgac 540
atccttcccc ttcttgagtg tgcttgcacc agcacctatg tgaatgagct tgtagtgttt 600
gttgttgtgg gcattgatat tggtgtgccc acagtcacca tcttcatttc ctatgctctc 660
attctctcca gcatcttcca cattgattcc acggagggca ggtccaaagc cttcagcacc 720
tgcagctccc acataattgc agtttctctg ttctttgggt caggagcatt catgtacctc 780
aaaccctttt ctcttttagc tatgaaccag ggcaaggtgt cttccctatt ctataccact 840
gtggtgccca tgctcaaccc attaatttat agcctgagga ataaggacgt caaagttgct 900
ctaaagaaaa tcttgaacaa aaatgcattc tcctga 936

<210> 37
<211> 945
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472443CB1

<400> 37
atggccaaga ataatctcac cagagtaacc gaattcattc tcatgggctt tatggaccac 60
cccaaattgg agattcccct ctttctggtg tttctgagtt tctacctagt caccttctt 120
gggaatgtgg ggatgattat gttaatccaa gtagatgtca aactctacac cccaatgtac 180
ttcttcctga gccacctctc cctgctggat gcctgttaca cctcagtcat caccctcag 240
atcctagcca cattggccac aggcaaaacg gtcatctcct acggccactg tgctgcccag 300
ttctttttat tcaccatctg tgcaggcaca gagtgctttc tgctggcagt gatggcctat 360
gatcgctatg ctgccattcg caacccactg ctctataccg tggccatgaa tcccaggctc 420
tgctggagcc tggtggtagg agcctatgtc tgtggggtgt caggagccat cctgcgtacc 480
acttgcacct tcaccctctc cttctgtaag gacaatcaaa taaacttctt cttctgtgac 540
ctcccacccc tgctgaagct tgcctgcagt gacacagcaa acatcgagat tgtcatcatc 600
ttctttggca attttgtgat tttggccaat gcctccgtca tcctgatttc ctatctgctc 660
atcatcaaga ccattttgaa agtgaagtct tcaggtggca gggccaagac tttctccaca 720
tgtgcctctc acatcactgc tgtggccctt ttctttggag cccttatctt catgtatctg 780
caaagtggct caggcaaatc tctggaggaa gacaaagtcg tgtctgtctt ctatacagtg 840
gtcatcccca tgctgaaccc tctgatctac agcttaagaa acaaagatgt aaaagacgcc 900
ttcagaaagg tcgctaggag actccaggtg tccctgagca tgtag 945

<210> 38
<211> 1041
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472445CB1

<400> 38
atgaatgcag ctcaaagcca ttatcctaag cgaactaatg caggaacagg aaaccaaata 60
tcacatgttc ttacttgtaa acaggcaaaa atatcaatgg gagaagaaaa ccaaaccttt 120
gtgtccaagt ttatcttcct gggtctttca caggacttgc agacccagat cctgctattt 180
atccttttcc tcatcattta tctgctgacc gtgcttggaa accagctcat catcattctc 240
atcttcctgg attctcgcct tcacactccc atgtattttt ttcttagaaa tctctccttt 300
gcagatctct gtttctctac tagcattgtc cctcaagtgt tggttcactt cttggtaaag 360
aggaaaacca tttcttttta tgggtgtatg acacagataa ttgtctttct tctggttggg 420
tgtacagagt gtgcgctgct ggcagtgatg tcctatgacc ggtatgtggc tgtctgcaag 480
cccctgtact actctaccat catgacacaa cgggtgtgtc tctggctgtc cttcaggtcc 540
tgggccagtg gggcactagt gtctttagta gataccagct ttactttcca tcttccctac 600
tggggacaga atataatcaa tcactacttt tgtgaacctc ctgccctcct gaagctggct 660
tccatagaca cttacagcac agaaatggcc atcttttcaa tgggcgtggt aatcctcctg 720
gcccctgtct ccctgattct tggttcttat tggaatatta tctccactgt tatccagatg 780
cagtctgggg aagggagact caaggctttt tccacctgtg gctcccatct tattgttgtt 840
gtcctcttct atgggtcagg aatattcacc tacatgcgac caaactccaa gactacaaaa 900
gaactggata aaatgatatc tgtgttctat acagcggtga ctccaatgtt gaaccccata 960
atttatagct tgaggaacaa agatgtcaaa ggggctctca ggaaactagt tgggagaaag 1020
tgcttctctc ataggcagtg a 1041

<210> 39
<211> 951
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472446CB1

<400> 39
atggagctct ggaacttcac cttgggaagt ggcttcattt tggtggggat tctgaatgac 60
agtgggtctc ctgaactgct ctgtgctaca attacaatcc tatacttgtt ggccctgatc 120
agcaatggcc tactgctcct ggctatcacc atggaagccc ggctccacat gcccatgtac 180
ctcctgcttg ggcagctctc tctcatggac ctcctgttca catctgttgt cactcccaag 240

gcccttgcgg actttctgcg cagagaaaac accatctcct ttggaggctg tgcccttcag 300
atgttcctgg cactgacaat gggtggtgct gaggacctcc tactggcctt catggcctat 360
gacaggtatg tggccatttg tcatcctctg acatacatga ccctcatgag ctcaagagcc 420
tgctggctca tggtggccac gtcctggatc ctggcatccc taagtgccct aatatatacc 480
gtgtatacca tgcactatcc cttctgcagg gcccaggaga tcaggcatct tctctgtgag 540
atcccacact tgctgaaggt ggcctgtgct gatacctcca gatatgagct catggtatat 600
gtgatgggtg tgaccttcct gattccctct cttgctgcta tactggcctc ctatacacaa 660
attctactca ctgtgctcca tatgccatca aatgagggga ggaagaaagc ccttgtcacc 720
tgctcttccc acctgactgt ggttgggatg ttctatggag ctgccacatt catgtatgtc 780
ttgcccagtt ccttccacag caccagacaa gacaacatca tctctgtttt ctacacaatt 840
gtcactccag ccctgaatcc actcatctac agcctgagga ataaggaggt catgcgggcc 900
ttgaggaggg tcctgggaaa atacatgctg ccagcacact ccacgctcta g 951

<210> 40
<211> 1395
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472451CB1

<400> 40
atggaaggca cattagattc tcctaatggc attatgcttt acaaatacct ggagagggat 60
gtgaacagca aggaactgca aagtggaaac cagacttctg tgtctcactt cattttggtg 120
ggcctgcacc acccaccaca gctgggagcg ccactcttct tagctttcct tgtcatctat 180
ctcctcactg tttctggaaa tgggctcatc atcctcactg tcttagtgga catccggctc 240
catcgtccca tgtgcttgtt cctgtgtcac ctctccttct tggacatgac catttcttgt 300
gctattgtcc ccaagatgct ggctggcttt ctcttgggta gtaggattat ctcctttggg 360
ggctgtgtaa tccaactatt ttctttccat ttcctgggct gtactgagtg cttcctttac 420
acactcatgg cttatgaccg tttccttgcc atttgtaagc ccttacacta tgctaccatc 480
atgacccaca gagtctgtaa ctccctggct ttaggcacct ggctgggagg gactatccat 540
tcacttttcc aaacaagttt tgtattccgg ctgcccttct gtggccccaa tcgggtcgac 600
tacatcttct gtgacattcc tgccatgctg cgtctagcct gcgccgatac ggccatcaac 660
gagctggtca cctttgcaga cattggcttc ctggccctca cctgcttcat gctcatcctc 720
acttcctatg gctatattgt agctgccatc ctgcgaattc cgtcagcaga tgggcgccgc 780
aatgccttct ccacttgtgc tgcccacctc actgttgtca ttgtttacta tgtgccctgc 840
accttcattt acctgcggcc ttgttcacag gagcccctgg atggggtggt agctgtcttt 900
tacactgtca tcactccctt gcttaactcc atcatctaca cactgtgcaa caaagaaatg 960
aaggcagcat tacagaggct agggggccac aaggaagaag tagaagaaat agagttgggt 1020
catacaactg tggaaggaca cagccttgcc acacagggac aacaaggtcc tcggcatttt 1080
gggcaccatg acagtgaaga accacaagtc aatgaaggac agattggtga ggaagtagta 1140
cttgggggtg tggaggtgag agtacaccct gatcaccagc aggatgagga ggttccccag 1200
cacaacttaa agaaaaatta tatgaccatc tcagcagatg gagaaaaagc actggacaat 1260
atccggcatc cattaataaa aactctcaat aacttagaag taaaagggga cttcctcaac 1320
ctgatgaagg atgtctatga aaatcctaca cctaacctaa gcaaatatcc tgaaaaacta 1380
aatgctttcc cctag 1395

<210> 41
<211> 972
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472456CB1

<400> 41
atgaaccctg aaaactggac tcaggtaaca agctttgtcc ttctgggttt ccccagtagc 60
cacctcatac agttcctggt gttcctgggg ttaatggtga cctacattgt aacagccaca 120
ggcaagctgc taattattgt gctcagctgg atagaccaac gcctgcacat acagatgtac 180
ttcttcctgc ggaatttctc cttcctggag ctgttgctgg taactgttgt ggttcccaag 240
atgcttgtcg tcatcctcac gggggatcac accatctcat ttgtcagctg catcatccag 300
tcctacctct acttctttct aggcaccact gacttcttcc tcttggccgt catgtctctg 360
gatcgttacc tggcaatctg ccgaccactc cgctatgaga ccctgatgaa tggccatgtc 420
tgttcccaac tagtgctggc ctcctggcta gctggattcc tctgggtcct ttgccccact 480
gtcctcatgg ccagcctgcc tttctgtggc cccaatggta ttgaccactt ctttcgtgac 540

agttggccct tgctcaggct ttcttgtggg gacacccacc tgctgaaact ggtggctttc 600
atgctctcta cgttggtgtt actgggctca ctggctctga cctcagtttc ctatgcctgc 660
attcttgcca ctgttctcag ggcccctaca gctgctgagc gaaggaaagc gttttccact 720
tgcgcctcgc atcttacagt ggtggtcatc atctatggca gttccatctt tctctacatt 780
cgtatgtcag aggctcagtc caaactgctc aacaaaggtg cctccgtcct gagctgcatc 840
atcacacccc tcttgaaccc attcatcttc actctccgca atgacaaggt gcagcaagca 900
ctgagagaag ccttggggtg gcccaggctc actgctgtga tgaaactgag ggtcacaagt 960
caaaggaaat ga 972

<210> 42
<211> 957
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472457CB1

<400> 42
atgaatccag caaatcattc ccaggtggca ggatttgttc tactggggct ctctcaggtt 60
tgggagcttc ggtttgtttt cttcactgtt ttctctgctg tgtattttat gactgtagtg 120
ggaaaccttc ttattgtggt catagtgacc tccgacccac acctgcacac aaccatgtat 180
tttctcttgg gcaatctttc tttcctggac ttttgctact cttccatcac agcacctagg 240
atgctggttg acttgctctc aggcaaccct accatttcct ttggtggatg cctgactcaa 300
ctcttcttct tccacttcat tggaggcatc aagatcttcc tgctgactgt catggcgtat 360
gaccgctaca ttgccatttc ccagcccctg cactacacgc tcattatgaa tcagactgtc 420
tgtgcactcc ttatggcagc ctcctgggtg gggggcttca tccactccat agtacagatt 480
gcattgacta tccagctgcc attctgtggg cctgacaagc tggacaactt ttattgtgat 540
gtgcctcagc tgatcaaatt ggcctgcaca gataccttg tcttagagct tttaatggtg 600
tctaacaatg gcctggtgac cctgatgtgt tttctggtgc ttctgggatc gtacacagca 660
ctgctagtca tgctccgaag ccactcacgg gagggccgca gcaaggccct gtctacctgt 720
gcctctcaca ttgctgtggt gaccttaatc tttgtgcctt gcatctacgt ctatacaagg 780
ccttttcgga cattccccat ggacaaggcc gtctctgtgc tatacacaat tgtcaccccc 840
atgctgaatc ctgccatcta taccctgaga aacaaggaag tgatcatggc catgaagaag 900
ctgtggagga ggaaaaagga ccctattggt cccctggagc acagaccctt acattag 957